



Recueil et intégration d'heuristiques dans les algorithmes de définition de stratégies de pompage

Pierre Lannuzel

► To cite this version:

Pierre Lannuzel. Recueil et intégration d'heuristiques dans les algorithmes de définition de stratégies de pompage. Interface homme-machine [cs.HC]. Ecole Nationale des Ponts et Chaussées, 1990. Français. NNT : . tel-00520786

HAL Id: tel-00520786

<https://pastel.archives-ouvertes.fr/tel-00520786>

Submitted on 24 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

X

Thèse de Doctorat

présentée à

l'Ecole Nationale des Ponts et Chaussées .

Spécialité

Sciences et techniques de l'Environnement.

par

Pierre LANNUZEL

Sujet de la thèse :

**Recueil et intégration d'heuristiques
dans les algorithmes de
définition de stratégies de pompage**

Thèse soutenue le 18 janvier 1990

Devant le jury composé de :

- M POCHAT, président,
- Me CORDIER, rapporteur,
- M LABYE, rapporteur,
- M GONDRAN, directeur de thèse,
- M LALEMENT,
- M SCHULHOF.

REMERCIEMENTS

L'auteur tient à remercier tous ceux qui ont contribué à la réalisation de ce travail de thèse, et plus particulièrement :

M GONDRAN, directeur de thèse, et l'ensemble du jury,

MM HAMON et DUBOIS, directeur et directeur adjoint de l'usine de Neuilly sur Marne, et l'ensemble des chefs de poste,

MM ARRAUDEAU, DAVOUST et GRUGET de la Compagnie Générale des Eaux,

M ORTOLANO, professeur à l'université de Stanford, dont les précieux conseils ont guidé la phase d'évaluation,

MM POCHAT et FREROT, directeur et directeur adjoint du CERGRENE,

Me PEROT, MM LEQUENNE et PENET, stagiaires ayant travaillé sur une partie de la thèse,

M ESTRANGIN et M CHAUVET, d'Euréquip et Neuron Data.

Le Syndicat des Eaux de l'Ile de France, et la Compagnie Générale des Eaux.

RESUME

PILOTE est un système qui reproduit des décisions d'experts dans la définition de stratégies de pompage. Pour cela, le système combine à la fois des modèles mathématiques et des jeux d'heuristiques. Les modèles comprennent des procédures statistiques de prévisions horaires de consommation et un module de simulation hydraulique. Les heuristiques sont basées sur des facteurs en relation avec la quantité et la qualité de l'eau, et les coûts de pompage et de maintenance. Une évaluation du système a indiqué que les procédures de prévision fonctionnaient bien, mais que le modèle de simulation surestimait de façon conséquente les volumes élevés. Les heuristiques utilisées dans le programme représentent bien les règles utilisées par les experts, excepté pour le classement des stratégies alternatives de pompage. L'évaluation a aussi défini les priorités de développement pour un passage vers un système opérationnel.

ABSTRACT

PILOTE is a computer program that attempts to reproduce decisions of expert operators in scheduling outlet pumps at a water treatment plant. The program integrates both mathematical models and heuristics (i.e., rules-of-thumb used by operators to select configurations of outlet pumps). The models include statistical procedures for predicting hourly water consumption and a hydraulic simulation for estimating hour-by-hour elevation changes at regulating reservoirs and at the elevated reservoir receiving the pumped flows. The heuristics used by operators are based on factors related to water quantity and quality and operation and maintenance costs. An evaluation of the program showed that the statistical forecasting procedures performed well, but the simulation model consistently overestimated the volumes pumped.

The heuristics included in the program accurately represented rules used by expert operators, except for rules used to rank order alternative pumping strategies. The evaluation provided an agenda of research for refining PILOTE so that it can be used routinely.

SOMMAIRE

INTRODUCTION	1 2
Ch1 PRESENTATION DU SUJET ET DE L'USINE	1 4
Ch2 L'APPORT DE L'UTILISATION DE CONNAISSANCES DANS LES TECHNIQUES DE RESOLUTION DE PROBLEMES.	4 6
Ch3 LE RECUEIL DE L'EXPERTISE	6 2
Ch4 DE L'INTEGRATION DE L'EXPERTISE A LA CONCEPTION DU SYSTEME	9 4
Ch5 L'IMPLEMENTATION DU LOGICIEL	1 1 9
Ch6 L'EVALUATION DU SYSTEME	1 4 0
CONCLUSION	1 8 2

Ch1	PRESENTATION DU SUJET ET DE L'USINE	1 4
I)	CONSIDERATIONS GENERALES SUR LES PROBLEMES SOLUBLES PAR SYSTEME-EXPERT.	
A)	Caractéristiques générales des problèmes solubles par un système-expert	
B)	Le choix d'une application.	
II)	LES BESOINS SUR L'USINE	1 7
A)	L'apparition de besoins d'optimisation de gestion	
B)	La recherche d'un problème illustratif	
III)	LA DESCRIPTION PHYSIQUE DE L'USINE	1 9
A)	Présentation générale	
B)	Les étapes de la filière	
IV)	L'ANALYSE FONCTIONNELLE DE L'USINE (aspect hydraulique)	2 3
A)	L'objet de cette analyse	
B)	Les variables de l'usine	
	1) Les variables de commande	
	2) Les variables d'état	
	3) La variable de consommation	
	4) Les autres variables	
	5) La formalisation sous forme d'automatisme	
	6) La nécessité de créer des modèles	
V)	LA CONSTRUCTION D'UN MODELE DE SIMULATION HYDRAULIQUE	2 8
A)	Les modèles initiaux sur l'usine	
B)	Le modèle de simulation développé	
	1) La modélisation du fonctionnement hydraulique	
	2) L'aspect énergétique de la simulation	

VI) LA MODELISATION DE LA CONSOMMATION

3 6

- A) La nécessité de construire un modèle
- B) L'analyse des données brutes
 - 1) Les données utilisées
 - 2) Les premières hypothèses formulées
 - 3) Des premières hypothèses à l'élaboration du modèle
- C) La construction des modèles "PROFIL"
 - 1) Le principe
 - 2) Caractéristiques du modèle par saison

VII) LE CADRAGE DU SUJET ET DE SES LIMITES

4 5

Ch2 L'APPORT DE L'UTILISATION DE CONNAISSANCES DANS LES TECHNIQUES DE RESOLUTION DE PROBLEMES.

4 6

I) L'INTRODUCTION DE LA NOTION D'HEURISTIQUE

- A) Les techniques d'exploration systématique de l'espace
 - 1) Les espaces d'état
 - 2) Exemples de problèmes définis dans les espaces d'état
- B) Recherche aveugle dans les espaces d'état
 - 1) Le formalisme habituel
 - 2) En largeur d'abord
 - 3) En profondeur d'abord
- C) Limites des recherches aveugles et introduction des heuristiques
- D) Exemples de fonctions d'évaluation
 - 1) Exemples dans le jeu du taquin
 - 2) Comparaison des décisions prises
- E) Complexité et efficacité des fonctions d'évaluation

II) RESOLUTION DE PROBLEMES EN UTILISANT DES HEURISTIQUES POUR RESTREINDRE L'ESPACE DE RECHERCHE

- A) L'élagage d'un arbre de recherche

B) Illustration par le programme DENDRAL (recherche d'une formule chimique)

- 1) Présentation générale
- 2) Les étapes de résolution du problème

III) L'INTERET D'UTILISER DES HEURISTIQUES DANS LE CAS DU PROBLEME POSE 6 0

A) La notion d'heuristique

B) La provenance des heuristiques

C) L'intérêt d'utiliser des heuristiques dans le cas traité

Ch3 LE RECUEIL DE L'EXPERTISE 6 2

I) CARACTERISTIQUES DES SCHEMAS DE RAISONNEMENT D'EXPERT

A) Quelques éléments sur l'ingénierie de la connaissance

B) Les acteurs d'un domaine de connaissance

- C) Points communs des raisonnements d'expert
- 1) Un raisonnement sous-tendu par un modèle cognitif
 - 2) Les caractéristiques d'un modèle d'expert

D) Deux notions centrales : le contexte et le schéma de décision

F) Les conséquences sur les caractéristiques d'un modèle d'expert

II) ORGANISATION D'UNE DEMARCHE DE RECUEIL D'EXPERTISE 6 9

A) Finalité de la démarche de recueil d'expertise

B) Identification et localisation de l'expertise du domaine

C) Le recueil des grains de connaissance

D) La construction du modèle cognitif de l'expert

E) Présentation de quelques techniques de recueil d'expertise

III) L'ACQUISITION DE L'EXPERTISE DE L'APPLICATION "PILOTE" 7 5

A) Les grandes lignes de la démarche suivie

B) La localisation de l'expertise

C) Les entretiens préliminaires avec le directeur technique

- 1) Les objectifs de cette phase
- 2) La définition des termes du domaine,
- 3) Les objectifs visés par l'expert
- 4) L'identification des principaux cas de figure
- 5) La structure d'un schéma décisionnel
- 6) Les critères d'évaluation des stratégies de conduite d'usine

D) L'observation sur le poste de commande

- 1) Les objectifs de cette phase d'observation
- 2) L'identification des informations structurantes des contextes
- 3) La mise à plat des schémas décisionnels
- 4) Les heuristiques de structuration de la journée
- 5) Les heuristiques de restriction des stratégies envisageables
- 6) Les heuristiques limitant le nombre de sous-pages
- 7) La prise en compte concrète des critères de choix des stratégies

D) L'élaboration de questionnaires

- 1) Les objectifs de cette phase
- 2) La recherche d'un consensus sur les découpages et points de repère
- 3) La recherche de consensus sur l'appréciation qualitative des critères

E) La confrontation et l'analyse des réponses

- 1) Le déroulement de la phase
- 2) Découpage de la journée et définition des objectifs
- 3) Importance et classes de qualification des critères

IV) CRITIQUE DE LA DEMARCHE DE PRISE D'EXPERTISE

9 1

A) Synthèse de la phase d'expertise

B) Les remarques issues de l'expérience

C) Expertise et système de traitement de l'information

D) Pertinence et accessibilité des informations

- 1) La pertinence d'une information
- 2) L'accessibilité d'une information
- 3) Une oscillation entre pertinence et accessibilité
- 4) Mise en place de protocoles de recherche de pertinence

Ch4 DE L'INTEGRATION DE L'EXPERTISE A LA CONCEPTION DU SYSTEME

9 4

I) LA CONSTRUCTION D'UNE METHODOLOGIE ADAPTEE AU PROBLEME

- A) Le passage du modèle de l'expert à la base de connaissance
- B) L'identification des caractéristiques du problème posé
- C) La démarche vue comme un système de traitement de l'information
 - 1) Les informations existantes
 - 2) Les informations issues d'un modèle
 - 3) Les informations issues d'un traitement
 - 4) L'utilisation de modèles dans leur ordre de complexité d'obtention
- D) Le principe de la méthodologie
 - 1) Un rapide rappel des méthodologies envisageables
 - 2) Les idées directrices de la démarche suivie

II) LA DEFINITION DU CONTEXTE ET LA DECOMPOSITION DU PROBLEME EN SOUS-PROBLEMES

102

- A) La nature de la décomposition
- B) Les règles de structuration de la journée
- C) Exemples de règles
- D) Les raisonnements à l'intérieur des périodes de planification

III) UNE PREMIERE SERIE DE FILTRAGES BASES SUR LES DONNEES INITIALES UTILISEES A PRIORI.

105

- A) La génération du premier noyau de stratégies potentielles
- B) Quelques exemples de règles
- C) L'efficacité de ces premiers filtrages

IV) UNE SECONDE SERIE DE FILTRAGES BASES SUR LES DONNEES QUANTITATIVES OBTENUES PAR UN MODELE SOMMAIRE.

106

- A) L'utilisation des points de repère avec introduction de la notion de fourchette

- B) L'efficacité de ce second niveau de filtrage.
- C) Du choix des chefs de poste à une démarche prenant en compte l'apport de l'informatique
- D) Une première étape vers l'apprentissage

V) UNE TROISIEME SERIE DE FILTRAGES BASES SUR LES DONNEES
QUANTITATIVES OBTENUES PAR LE MODELE DE SIMULATION. 109

- A) L'introduction d'un second niveau dans le raisonnement
- B) Les informations apportées par le modèle de simulation
 - 1) Une description quantitative sur le déroulement prévisionnel de la plage horaire.
 - 2) Une meilleure précision sur les états prévisionnels de fin de plage
- C) L'efficacité de ce niveau de sélection numérique
- D) Le traitement des informations issues de la simulation

VI) LA SELECTION D'UNE STRATEGIE PAR UN RAISONNEMENT QUALITATIF

- A) La formulation du problème 111
- B) L'extraction des informations pertinentes
- C) Le regroupement des critères en niveaux de qualification
- D) La présentation de deux méthodologies envisageables
- E) Limites de ce type d'approche
- F) La méthodologie retenue : une sélection lexicographique par critères

I) L'IMPLEMENTATION D'UNE METHODOLOGIE

- A) Le problème posé
- B) Critères de choix d'outils de développement
 - 1) Choisir un langage ou un générateur
 - 2) Le formalisme de représentation des connaissances
 - 3) Les stratégies de résolution potentielles de l'outil
 - 4) La visualisation du raisonnement
 - 5) Le degré de finition de l'outil
 - 6) Les supports de développement
 - 7) Autres considérations sur le choix
- C) Les deux configurations successives retenues
 - 1) La première option
 - 2) La seconde option

II) LA STRUCTURE DE LA MAQUETTE PILOTE

- A) L'architecture de la maquette
 - 1) Une décomposition en modules
 - 2) Les bases de connaissances et les programmes C
 - 3) Les passages de valeurs entre bases et programmes
 - 4) Le schéma de la maquette
 - 5) L'algorithme du raisonnement
- B) Description du programme PREVISION
 - 1) Fonctionnalités du programme
 - 2) Les entrées
 - 3) Les sorties
- C) Description de la base "CONTEXTE-CRITERES"
 - 1) Les fonctionnalités
 - 2) L'organisation de la base
 - 3) Les entrées et sorties de la base
- D) La définition de l'hypothèse de régime nourricier
 - 1) Les fonctionnalités
 - 2) Les entrées et sorties de la base
 - 3) L'organisation de la base
- E) Description des modules du programme "PILOTE"
 - 1) Les fonctionnalités

- 2) Le module de génération des stratégies
- 3) Le module de simulation
- 4) Le module de sélection
- 5) Le regroupement des plages horaires
- 6) Le module de classement des plages horaires

F) Les fichiers de PILOTE

Ch 6 L'EVALUATION DU SYSTEME

1 4 0

I) LE ROLE ET L'ORGANISATION DE CETTE PHASE

A) Une évaluation et non une validation

- 1) L'organisation de la phase d'évaluation
- 2) Une première évaluation module par module
- 3) L'évaluation de l'enchaînement des modules et des plages horaires

II) L'EVALUATION DES MODELES

1 4 2

A) L'évaluation du modèle de prévision de consommation

- 1) Les deux aspects de l'évaluation
- 2) Les modalités de l'étude
- 3) L'évaluation de la précision de la prévision
- 4) Incidence des erreurs de prévision sur les décisions

B) L'évaluation du modèle de simulation

- 1) Le principe de l'évaluation de ce modèle
- 2) Les résultats

III) L'EVALUATION DES HEURISTIQUES DES BASES DE CONNAISSANCES

1 5 8

A) Le sens et la méthodologie de l'évaluation

B) La génération des stratégies

C) Les heuristiques du premier niveau de filtrage

D) Les heuristiques de sélection des configuration

IV) L'EVALUATION GLOBALE

1 6 2

A) L'évaluation des enchaînements

- 1) L'aspect technique
- 2) Une comparaison difficile avec les experts

B) L'évaluation des objectifs définis lors de la conception	
1) La modularité du système	
2) La mémorisation de la trace du raisonnement	
3) L'efficacité de l'utilisation des informations	
C) Une réflexion sur le renforcement des heuristiques de sélection	
1) Une orientation de la réflexion vers l'agrégation de critères	
2) Présentation de la démarche envisagée	
3) La construction de l'arbre de regroupement	
4) La définition des classes de qualification	
5) Méthodologie d'utilisation de l'agrégation	
6) Intérêt de l'agrégation	
D) Une évaluation des gains potentiels	
V) UNE ORIENTATION VERS LES PROGRAMMES D'APPRENTISSAGE	1 7 1
A) L'objet de l'apprentissage dans PILOTE	
B) Eléments caractérisant les programmes d'apprentissage	
1) Définition de l'apprentissage	
2) Finalités des principaux programmes	
3) Raisonnement inductif et déductif	
4) Eléments d'un programme d'apprentissage	
5) Classification des programmes	
VI) CONCLUSION DE LA PHASE D'EVALUATION	1 8 1
CONCLUSION GENERALE	1 8 2
BIBLIOGRAPHIE	

ANNEXE 1 : Les caractéristiques de quelques outils de développement de systèmes experts

- A) Les langages de manipulation symbolique
- B) Un générateur dans un logiciel intégré : GURU
- C) Un générateur basé sur la gestion d'objets : NEXPERT-Objet
- D) Un générateur basé sur la logique des prédicats d'ordre 1 : l'Expert-Kit
- E) Un générateur basé sur la logique des prédicats d'ordre 2 : SNARK

ANNEXE 2 : Présentation de différents programmes d'apprentissage

- I) Apprentissage par réorganisation de la connaissance
- II) Apprentissage par instruction
- III) Apprentissage déductif par recherche d'explications
- IV) Apprentissage inductif à partir d'exemples
- V) Apprentissage par expérimentation
- VI) Apprentissage inductif à partir d'observations
- VII) Apprentissage par analogie.

**ANNEXE 3 Le texte de l'article paru dans le numéro de juillet du "Journal of Water Resources Planning and Management".
Revue de l'ASCE.**

INTRODUCTION

Le sujet de la thèse vient d'une préoccupation commune à une grande entreprise intervenant dans le domaine de l'eau, et à un centre de recherche dont le champ d'activité recouvre les sciences et techniques liées à la gestion de l'eau.

Dans les années 1960, les préoccupations des compagnies productrices et distributrices d'eau étaient principalement la réalisation d'investissements lourds. Dans la dernière décennie, les progrès dans le domaine de l'eau ont été essentiellement obtenus par une optimisation de la gestion, c'est à dire par la définition de stratégies de pompage dans le cas de l'eau potable, ou de mouvements de vannes dans les réseaux d'assainissement.

Jusqu'à ces dernières années, la définition de ces stratégies faisait essentiellement appel aux techniques de la recherche opérationnelle, et utilisaient des procédures de calcul essentiellement algorithmiques.

Les progrès en intelligence artificielle ont orienté, il y a 3 ans, lors de la définition du sujet de thèse, vers une autre approche méthodologique, s'appuyant sur l'intégration de l'expérience concrète accumulée par ceux qui avaient la charge de définir de telles stratégies. En effet, le développement des systèmes experts a mis en évidence l'intérêt d'utiliser l'expérience acquise par les experts du domaine, ou d'une manière plus générale, par les exploitants. L'intégration d'heuristiques, acquises par des années de pratique, peut affiner dans pratiquement tous les cas, les processus de recherche de solutions.

La ligne directrice de cette thèse, en sciences et techniques de l'environnement, s'inscrit donc dans une démarche d'extraction, puis d'intégration d'une expertise existante dans la définition d'une méthodologie de recherche de stratégies.

Le premier point central du travail de thèse a donc été le recueil des heuristiques auprès des experts. Collecter une expertise, signifie définir une approche vis à vis des chefs de poste, contrôler et analyser avec une vue critique, les éléments recueillis. Comme, dans ce cas, l'expertise était partagée entre plusieurs experts, une homogénéisation des pratiques a été nécessaire.

Une décision de choix de stratégie de pompage, fait intervenir beaucoup de paramètres et ne peut se prendre uniquement à partir d'un jeu d'heuristiques. Des précisions indispensables ne s'obtiennent que par l'utilisation de modèles numériques, en particulier pour les prévisions et la simulation. Une seconde partie du travail de thèse a donc consisté à

construire ces modèles, et à les combiner avec le jeu d'heuristiques recueillies pour définir une méthodologie de recherche de stratégies de pompage.

Le système a donc été conçu pour combiner un modèle qualitatif avec des modèles numériques, avec comme ligne directrice la volonté d'utiliser, à chaque étape du raisonnement, le niveau d'information qui réduise, au mieux et à moindre coût de traitement, l'espace de recherche. Ceci a conduit à définir une méthode basée sur des cycles de génération, test et sélection de stratégies.

Cette méthodologie a été illustrée par la construction d'une maquette. Pour le choix d'un outil de développement, plusieurs solutions étaient envisageables. Aussi, les considérations sur les choix effectués ont été détaillés, en donnant d'une part les caractéristiques des outils (en annexe), et d'autre part, les arguments de la solution retenue. La maquette est alors décrite, telle qu'elle a été construite et présentée, aux responsables de la CGE et du SEDIF, et au personnel concerné de l'usine.

Enfin, une phase d'évaluation a été menée en collaboration avec le Professeur ORTOLANO, de l'Université de Stanford, Californie. D'abord, les modèles et les heuristiques ont été évalués séparément. Puis les enchaînements des modèles qualitatifs et numériques ont été testés. L'évaluation a indiqué les priorités de développement pour la suite du projet, avec, en particulier, le renforcement du modèle de simulation.

Cette phase a aussi induit une orientation de la réflexion dans deux axes. Le premier a pour objet la recherche de nouveaux schémas d'utilisation des critères de choix des stratégies de pompage. Le second découle du fait que l'utilisation concrète de ce type de méthodologie génère une expertise, intéressante à intégrer aux sessions suivantes, par le fait que les heuristiques initiales peuvent être affinées par l'intermédiaire des modèles numériques. Ceci a orienté la dernière partie de la thèse vers les notions d'apprentissage automatique. Dans le cadre de cette thèse, le problème de l'apprentissage a été simplement abordé sous une forme de recherche bibliographique.

Ch 1 : PRESENTATION DU SUJET ET DE L'USINE

I) CONSIDERATIONS GENERALES SUR LES SYSTEMES EXPERTS

A) Caractéristiques générales des problèmes solubles par système-expert

Un problème est défini par un écart entre un état initial, tel qu'il est perçu, et un état final, tel qu'il est souhaité.

Newell et Simon ont postulé que la résolution de problèmes prenait place, au sein d'un système de traitement de l'information, dans un espace de représentation et mettait en oeuvre des méthodes de recherche générales et spécifiques.

Pour eux, un problème d'intelligence artificielle a comme caractéristique qu'il n'y a pas de manière unique de construire l'espace de représentation, ni de mener la résolution dans cet espace, ce point illustrant la capacité des individus à s'adapter à l'environnement d'une tâche. Un processus de compréhension intègre donc divers types de connaissances dans un tout cohérent, formant une construction mentale. Cette construction comprend l'espace de représentation et l'ensemble des méthodes et connaissances activées par la recherche d'une solution dans cet espace.

Plusieurs ouvrages [BONNET 86], [HAYES-ROTH 83] font référence à une typologie de problèmes abordables par la technique des systèmes experts.

La liste suivante n'est pas exhaustive, mais est représentative des principaux types de problèmes :

- le diagnostic ponctuel dans le temps,
- l'ordonnancement des tâches et l'affectation de ressources,
- la sélection d'informations pertinentes,
- la conception d'objets,
- l'aide à la conceptualisation d'un domaine,
- l'interprétation de situations,
- la prédiction,
- le "débuggage".

B) Le choix d'une application

Toujours d'après Bonnet, le choix d'une application doit résulter d'un processus qui garantisse à la fois, quatre critères principaux, qui sont :

- l'exemplarité de l'expérience,
- la participation active et la motivation des utilisateurs potentiels, des experts et des dirigeants,
- la prise en compte des objectifs stratégiques de l'entreprise,
- des chances raisonnables de réussite finale du projet.

1) L'exemplarité de l'expérience

Pour être démonstratif, le sujet retenu doit être simple, mais non trivial. Il doit présenter un degré de généralité suffisant pour qu'il soit transposable sur d'autres applications.

Il doit aussi y avoir une différence significative entre les performances de l'expert et celles d'un individu quelconque.

Une partie des informations manipulées doit être de nature qualitative, ou des décisions doivent être prises en présence d'informations incomplètes.

2) La participation et motivation des personnels

Les dirigeants et les futurs utilisateurs doivent être convaincus de l'utilité d'un système, et il doit exister au moins un expert susceptible de collaborer au projet, et d'y consacrer régulièrement du temps.

3) La prise en compte des objectifs stratégiques de l'entreprise

Plusieurs cas de figure peuvent se présenter.

Des connaissances utiles à l'entreprise ne sont ni consignées dans des manuels, ni enseignées, et risquent donc une disparition de façon impromptue.

L'expertise est partagée entre plusieurs personnes alors qu'une homogénéisation des pratiques ou des décisions est nécessaire au bon fonctionnement de l'entreprise.

L'expertise est concentrée en un seul endroit, alors que les utilisateurs sont dispersés.

La diffusion de l'expertise entraîne des coûts élevés pour l'entreprise.

4) Des chances réelles de réussite

Ceci s'apprécie d'une part, en évaluant les moyens matériels et en personnel disponibles, face à l'ampleur du problème, et d'autre part en se référant à certaines caractéristiques du problème à traiter.

Newell et Simon [NEWELL, SIMON 69] ont apporté un certain nombre d'idées sur la complexité des problèmes gérables par un système expert.

Tout d'abord, le problème doit être bien défini, c'est à dire que l'on doit pouvoir donner une bonne description de l'état initial et de l'état final du problème. Dans le cas d'un problème de planification, où interviennent des aléas, l'état final ne peut pas être décrit, ce sont donc des objectifs précis qu'il importe de fixer.

D'autres constantes sont décrites par Newell et Simon. Ils indiquent ainsi que le nombre de symboles manipulés est de l'ordre de quelques dizaines, au plus une centaine. De même, le processus de résolution demande une à deux dizaines de minutes, et le nombre d'hypothèses gérées est faible, et la remise en cause de ces hypothèses est rare.

L'espace de problème peut être ainsi vu comme un espace clos, comprenant l'ensemble des états possibles des connaissances utiles à la résolution (objets, relations entre objets, informations gérant le processus de recherche).

II) LES BESOINS SUR L'USINE

A) L'apparition de besoins d'optimisation de gestion

Depuis plusieurs années, les plans d'équipement dans les usines de production d'eau ont surtout été destinés à la mise en place de systèmes de collecte et de stockage des données.

Ainsi, sur l'usine de Neuilly, venait d'être installé un système de mesure des pressions, et de calcul des débits des pompes. Ces mesures, prises par capteurs avec un pas de temps très fin, sont archivées par l'UCO (unité centrale ordinateur), sur bandes magnétiques.

De même, le relevé sur papier des différents niveaux de réservoirs a été remplacé par une saisie informatique (programme SOPHIA), ce qui a constitué des fichiers de données pour traiter les prévisions de consommations.

A partir du moment où était mis en place ces systèmes de collecte de données, une phase d'utilisation de ces données pouvait être envisagée.

Collecter des informations représente une charge importante, tant du point de vue de l'investissement en matériel, que de celui de la saisie quotidienne des données sur l'usine. Or tout investissement appelle des retours, et à l'ordre du jour de la vie de l'usine, se posait la question de l'optimisation de la gestion à partir d'un traitement des informations collectées ou saisies.

Au moment où se discutait le sujet de la thèse, une des préoccupations de l'usine était la planification de la gestion hydraulique et chimique de la production d'eau.

B) La recherche d'un problème illustratif

Si le choix d'un problème de planification avait été arrêté d'emblée, il restait encore à délimiter le sujet, pour qu'il soit abordable par une équipe extrêmement restreinte, dont on savait dès le départ qu'elle ne serait formée que d'un élève-chercheur, aidé ponctuellement par quelques stagiaires, et ne pouvant faire appel qu'à une disponibilité limitée du personnel de l'usine.

Les problèmes de planification sont nombreux dans une usine de production d'eau, et portent aussi bien sur la planification du lavage des filtres, que sur les configurations de pompage, ou encore sur les quantités de produits chimiques à injecter lors du traitement de l'eau.

Le problème de l'optimisation du traitement chimique de l'eau constituait un sujet intéressant, mais pour lequel se posait un problème d'encadrement au niveau de l'expertise. Par contre pour la gestion hydraulique on disposait d'experts prêts à collaborer, avec en particulier, le directeur technique de l'usine, et plusieurs chefs de poste.

Le cadre d'une thèse veut aussi un certain niveau de généralité dans la démarche de recherche et les conclusions. De ce point de vue aussi, la recherche de configurations de pompage apparaissait intéressante à développer sur le plan méthodologique, car abordée jusqu'alors, principalement par des calculs de programmation dynamique.

Tous ces éléments ont orienté le choix du sujet de la thèse vers la recherche de configurations de pompage. Ce sujet correspond à une préoccupation réelle de la vie de l'usine. Il peut être illustratif des méthodes de raisonnement utilisées en intelligence artificielle, et présenter une voie alternative à la recherche de solutions purement algorithmique.

Donner de réelles chances de succès au projet demandait de bien délimiter les fonctionnalités attendues. Ce point est abordé dans le dernier paragraphe de ce chapitre, car les limites données au problème ne sont compréhensibles qu'une fois l'usine décrite et son fonctionnement analysé.

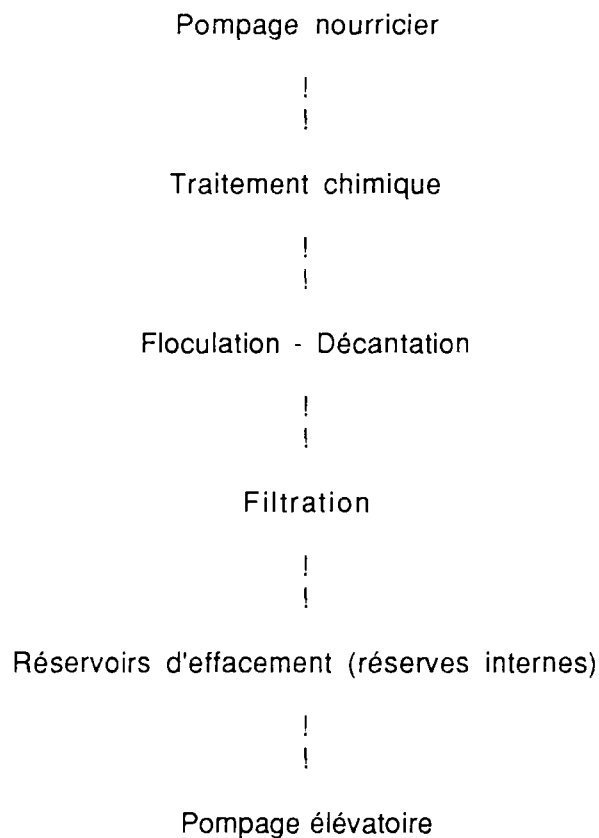
III) LA DESCRIPTION PHYSIQUE DE L'USINE

A) Présentation

L'usine de Neuilly-sur-Marne est l'une des trois usines de production d'eau gérées par le Syndicat des Eaux de l'Ile de France. Elle est implantée, de part et d'autre de la Marne, sur les territoires des communes de Neuilly-sur-Marne, Noisy-le-Grand et sa capacité de production est de 600 000 m³/jour.

B) Les étapes de la filière

Le schéma ci-dessous représente les différentes phases de la production d'eau :



1) L'usine nourricière :

Elle prélève l'eau dans la Marne et est composée de 6 groupes électro-pompes :

- 4 pompes BERGERON type SV 175 à débit variable entre 4000 et 7000 m³/heure.

- 2 pompes BERGERON type SX 238 d'un débit unitaire de 12 450 m³ par heure sous une hauteur manométrique totale de 14 mètres.

2) La mise en oeuvre des réactifs :

La cuve de traitement de l'eau brute comporte 3 files de traitement qui chacune peut admettre un débit compris entre 130 000 et 400 000 m³ /jour. Le principal réactif injecté à ce niveau est le WAC qui peut être introduit à raison de 50 g/m³.

3) La floculation-décantation :

Après avoir été brassée dans les flocculateurs, l'eau circule ensuite dans des décanteurs de type couloir à étages durant un minimum de 2 heures.

4) La filtration :

Une fois décantée, l'eau subit une filtration sur sable. Chaque filtre est piloté par une hotte à dépression de type polhydra, système de régulation qui permet d'assurer un débit égal pour chacun des bassins et une vitesse constante de filtration, quel que soit le colmatage du lit filtrant.

Les filtres, lorsqu'ils sont colmatés, sont lavés automatiquement par un flux d'eau et d'air établi à contre-courant.

5) L'affinage et la stérilisation :

L'eau clarifiée subit, en fin de filière, un traitement d'affinage assurant sa stérilisation. Ce traitement est réalisé par l'ozone.

6) Les réservoirs d'effacement :

Situés juste en amont de l'usine élévatoire, ils servent de tampon entre celle-ci et l'usine nourricière. Ces trois réservoirs ont les caractéristiques suivantes :

- R1 : Capacité : 19 000 m³
Cotes : radier : 36,20; trop-plein : 41 m.
- R2 : Capacité : 15 000 m³
Cotes : radier : 36,30; trop-plein : 41 m.
- R3 : Capacité : 27 500 m³
Cotes : radier : 34,50; trop-plein : 41 m.

7) L'usine élévatoire :

Elle est formée de l'ancienne et de la nouvelle usine :

L'ancienne usine, aspirant l'eau des réservoirs R1 et R2, comporte 8 groupes électropompes :

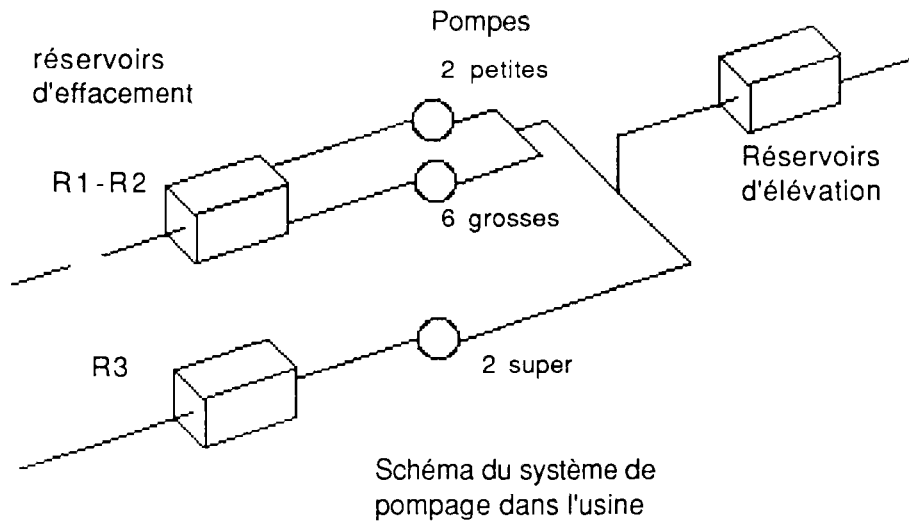
- 2 pompes BERGERON type SS 63 d'un débit unitaire de 1250 m³ par heure sous une hauteur manométrique totale de 110 mètres. Ces pompes sont désignées ultérieurement par e11-e12.

- 6 pompes BERGERON type SX 238 d'un débit unitaire de 4166 m³ par heure sous une hauteur manométrique totale de 112 mètres. Ces pompes sont désignées ultérieurement par E1-E6.

La nouvelle usine, aspirant l'eau du réservoir R3, comporte 2 groupes électropompes:

- 2 pompes Alsthom Atlantique Rateau d'un débit unitaire de 11 250 m³ par heure sous une hauteur manométrique totale de 90 mètres. Ces pompes sont désignées ultérieurement par E7-E8.

Le schéma du système est donné ci-dessous.



8) Les réservoirs de première élévation :

Ils sont au nombre de deux : Montreuil et Gagny.

Montreuil est formé de deux groupes de réservoirs. Le groupe 1 est formé du bassin n°6 de cote maximale 4,4 mètres, où 1 m. d'eau correspond à 4900 m³. Le groupe 2 est formé des réservoirs 7, 8, 9 battant ensemble, 1 m d'eau correspondant à 20 000 m³.

Pour des raisons de qualité d'eau, le réservoir 6 est vidangé une fois par semaine en hiver, et deux fois en été au moyen de l'usine de deuxième élévation de Montreuil.

Les quatre réservoirs de Gagny sont reliés entre eux; 1 m d'eau correspondant à 5300 m³.

IV) L'ANALYSE FONCTIONNELLE DE L'USINE

A) L'objet de cette analyse

L'objet de cette phase est de définir les variables de commande, les variables d'état, et les relations entre ces variables.

B) Les variables de l'usine :

L'usine, telle qu'elle est décrite dans le chapitre précédent se modélise à partir d'un jeu de 3 variables de commandes et 4 d'état.

1/ Les variables de commande :

La première décision consiste à définir un régime nourricier. Cette variable est en fait pratiquement continue.

La seconde décision concerne la répartition des volumes d'eau filtrée vers les réservoirs R_1 - R_2 alimentant l'ancienne élévatoire, et vers le réservoir R_3 alimentant la nouvelle élévatoire. Cette variable a une valeur comprise entre 0, quand toute l'eau est déviée vers la nouvelle élévatoire et 1 quand elle est au contraire déviée vers l'ancienne.

La troisième variable de décision correspond au choix d'une configuration élévatoire. Celle-ci étant composée de pompes à vitesse fixe, la connaissance du nombre de groupes de chaque type en fonctionnement, lié aux variables d'état, permet de calculer le débit refoulé.

Cette variable se présentera donc sous la forme d'un vecteur (i, j, k). Les nombres i, j, k désignent le nombre de pompes de chaque type en fonctionnement :

i : pour le type $E_7 - E_8$ i compris entre 0 et 2

j : pour le type $E_1 - E_6$ j compris entre 0 et 6

k : pour le type $E_{11} - E_{12}$ k compris entre 0 et 2

2/ Les variables d'état :

a/ Le niveau du réservoir de Montreuil :

L'évolution de ce niveau est liée d'une part à la fonction décrivant la consommation en eau du réseau, et d'autre part à l'apport élévatoire, dépendant de la troisième variable de commande. Le réservoir de Gagny, ouvert seulement quelques heures pendant la nuit, est considéré comme un consommateur.

b/ Le niveau du réservoir R1 :

Les variations de ce niveau dépendent directement des 2 premières variables de commande, définissant le régime nourricier, et le coefficient de répartition, et du débit refoulé par l'élévatoire.

c/ Le niveau du réservoir R3 :

Cette dernière variable d'état dépend de la 3ème variable de commande, de la première variable d'état, et de la fonction décrivant la consommation en eau du réseau.

3) La variable de consommation :

Cette variable décrit sous forme d'un tableau, le volume que doit produire l'usine, heure par heure, durant toute la période de planification.

4) Les variables dépendantes :

Les variables ci-dessous décrivant certains aspects du système, sont calculables à partir des variables d'état, de commande, et de la consommation.

a/ La pression de refoulement :

Cette variable représente la pression à l'aval des pompes élévatoires. Elle est une fonction linéaire de la demande, du niveau du réservoir de Montreuil et dépend de la configuration de pompage.

b/ Les hauteurs manométriques totales :

La hauteur manométrique totale représente la différence de pression entre l'aval et l'amont d'une pompe et sa connaissance permet de calculer le débit refoulé à partir des équations des courbes spécifiques aux pompes. Les hauteurs manométriques de l'ancienne et de la nouvelle élévatoire dépendent linéairement de la pression de refoulement et respectivement des niveaux des réservoirs R1-R2 et R3.

c/ L'énergie électrique consommée :

La consommation d'énergie électrique pour le pompage se calcule à partir d'une équation du second degré dont la variable est la hauteur manométrique et dont les coefficients dépendent de la configuration, c'est à dire de la variable de commande.

5) La formalisation sous forme d'automatisme

LES VARIABLES DE COMMANDE

U_{MON} : volume horaire pompé par la nourricière $10\,000 < U_{MON} < 25\,000$

U_{cp} : coefficient de répartition de l'eau entre les réservoirs R_1 et R_3 $0 \leq U_{cp} \leq 1$

Up : configuration de pompage, défini sous forme d'un vecteur à 3 dimensions (i, j, k)

i : nombre de pompes de la nouvelle élévatoire en route $0 \leq i \leq 2$

j : nombre de grosses pompes de l'ancienne élévatoire en route $0 \leq j \leq 6$

k : nombre de petites pompes de l'ancienne élévatoire en route $0 \leq k \leq 2$

LES VARIABLES D'ETAT

X_{R1} et X_{R3} : niveaux des réservoirs internes R_1 et R_3

X_{MON} : niveau du réservoir de Montreuil

X_{PREF} : pression de refoulement dans le réseau.

LA VARIABLE CONSOMMATION

Conso. : prévision horaire de consommation.

LES CONTRAINTES DU SYSTEME

Les limites physiques des réservoirs

$$0,8 \text{ m} \leq X_{R1} \leq 4 \text{ m}$$

$$0,8 \text{ m} \leq X_{R3} \leq 6 \text{ m}$$

$$3,5 \text{ m} \leq X_{mon} \leq 9 \text{ m}$$

Les pressions de refoulement

$$X_{PREF} < 105 \text{ m}$$

La puissance électrique consommée

$$P_{conso} < P_{disponible}$$

$P_{disponible}$ dépend de l'heure, du jour et de la saison

LES RELATIONS ENTRE LES VARIABLES

1) L'évolution du niveau du réservoir R₁

$$X_{R1}(t) = \frac{(U_{MON} * U_{CP} - \text{volume } 1)}{\text{surf. } R1} + X_{R1}(t-1) \quad (F1)$$

$$\text{avec volume } 1 = j (d + a * h_{mt1} + b * h_{mt1}^2) + k (d' + a' * h_{mt1} + b' * h_{mt1}^2) \quad (F2)$$

$$\text{et } h_{mt1} = X_{PREF} + C_1 - X_{R1}(t-1) \quad (F3)$$

2) L'évolution du niveau de réservoir R₃

$$X_{R3}(t) = \frac{(U_{MON} * (1 - U_{CP}) - \text{volume } 2)}{\text{surf. } R3} + X_{R3}(t-1) \quad (F4)$$

$$\text{avec volume } 2 = i (d_2 + a_2 * h_{mt2} + b_2 * h_{mt2}^2) \quad (F5)$$

$$\text{et } h_{mt2} = X_{PREF} + C_2 - X_{R3}(t-1) \quad (F6)$$

3) L'évolution du réservoir de Montreuil

$$X_{MON}(t) = \frac{(\text{volume } 1 + \text{volume } 2 - \text{conso.})}{\text{surf. } MON} + X_{MON}(t-1) \quad (F7)$$

4) La pression de refoulement

$$X_{PREF} = C^{te} + f(i, j, k) + k_0 * \text{conso.} + X_{MON}(t-1) \quad (F8)$$

LE CALCUL DE LA PUISSANCE CONSOMMEE

$$\begin{aligned} P_{\text{conso.}} = & i (k_1 + k_2 * h_{mt2} + k_3 * h_{mt2}^2) + j (K'_1 + k'_2 * h_{mt1} + k'_3 * h_{mt1}^2) \\ & + k (k''_1 + k''_2 * h_{mt1} + k''_2 * h_{mt1}^2) \end{aligned} \quad (F9)$$

6) Une possibilité et une nécessité : la création de modèles

L'analyse fait apparaître à la fois la nécessité et la possibilité de construire un modèle de simulation de l'usine, et un modèle prévisionnel de consommation d'eau. En effet, toute planification est impossible sans l'utilisation d'un modèle de prévision, aussi sommaire soit il. Dans le cadre du développement méthodologique, il a donc fallu construire ces modèles, en s'appuyant sur ce qui pouvait exister à l'usine.

V) LA CONSTRUCTION D'UN MODELE DE SIMULATION

A) Les modèles initiaux sur l'usine

Lors de la définition du sujet de la thèse, l'usine disposait de deux modèles de simulation hydraulique.

Le premier est un modèle sommaire, développé et utilisé par les chefs de poste. Ce modèle associe à chaque pompe un débit moyen et une puissance consommée moyenne, indépendamment de la hauteur manométrique.

Ce modèle sommaire ne donne que l'ordre de grandeur de ce qui sera élevé dans un intervalle de temps par une configuration de pompage.

La Compagnie Générale des Eaux, dispose en son centre d'un modèle sophistiqué, appelé PROGRES, pouvant donner débit et pression en tout point du réseau, à partir d'un calcul complet.

Ce modèle, n'a pas été utilisé, et n'est donc pas décrit, parce qu'il demande pour produire des résultats, un nombre important de données, incompatible avec le développement d'une maquette.

Il est vite apparu que le modèle sommaire l'était vraiment trop, et comme PROGRES n'était pas utilisable dans le contexte de l'étude, il a été nécessaire d'élaborer un modèle de simulation hydraulique, le plus performant possible, mais ne faisant appel qu'aux informations disponibles sur l'usine.

B) Le modèle de simulation développé

Le développement du modèle de simulation a été basé sur la documentation disponible à l'usine (ex. les courbes de pompes), sur l'étude des pressions de refoulement faite par M LEQUENNE en juin 87, et enfin sur l'analyse d'informations obtenues à partir d'enregistrements sur l'UCO (puissances consommées, débits).

Le modèle présenté dans ce document a essayé d'intégrer l'ensemble de la connaissance disponible en ce moment. Cette connaissance étant parcellaire, le développement du modèle a nécessité des hypothèses simplificatrices qui en réduisent l'exactitude.

La principale simplification concerne le calcul de la pression de refoulement. Elle fait suite à une étude statistique [LEQUENNE 87] et n'est pas basée sur un calcul complet de réseau. Cette approche n'est acceptable que pour les situations répertoriées dans l'étude statistique. Cette hypothèse et

les autres sont présentées dans ce chapitre et ont été testées pendant la phase d'évaluation décrite dans un chapitre suivant.

1) Modélisation du fonctionnement hydraulique sur l'usine :

a) Les fonctionnalités et l'organigramme du modèle :

Le modèle de simulation a pour fonction de déterminer, heure par heure, et pendant la durée de la plage horaire l'évolution des variables d'état du système, c'est à dire :

- les niveaux prévisionnels des réservoirs d'effacement,
- le niveau prévisionnel du réservoir de Montreuil.

Le modèle calcule également la pression de refoulement et la puissance électrique potentiellement appelée à chaque pas de temps.

b) Les données initiales

Le modèle nécessite la connaissance des données initiales suivantes :

- Hauteurs manométriques initiales (hmt),
- Niveaux des réservoirs d'effacement et de Montreuil,
- Hypothèse de régime nourricier établie par heure sur 24 heures,
- Demande prévisionnelle en eau établie par heure sur 24 heures.

c) La boucle de calcul (avec en référence les formules utilisées)

Configuration + hmt (i)
!
(F2 et F5)
Volumes élevés par les élévatoires (i-i+1)
!
(F9)
Puissance électrique appelée (i-i+1)
!
(modèle prévisionnel)
Prévision de consommation
!
Régime nourricier
!
(F1 F4 et F7)
Evolutions de niveau des réservoirs (i-i+1)
!
(F8)
Pression de refoulement (i+1)
!
(F3 et F6)
hmt (i+1)

d) Le calcul des volumes refoulés par l'élévatoire :

i) Les relations Débit-HMT :

Toutes les pompes d'une même famille sont caractérisées par une courbe "débit-HMT", qui peut être formalisée par une relation mathématique du type :

$$D = D^0 + a.HMT - b.HMT^2$$

Ce débit est ensuite multiplié par un coefficient correcteur cr.

a) Pour les pompes E1-E6, les coefficients de l'équation sont :

$$D^{\circ}_1 = - 4160 ; \quad a_1 = 280 ; \quad b_1 = - 1,91 ;$$

Le coefficient correcteur est $cr_1 = 0,96$.

b) Pour les pompes e11-e12, ces coefficients sont :

$$D^{\circ}_2 = 360 ; \quad a_2 = 47,9 ; \quad b_2 = - 0,351 ;$$

Le coefficient correcteur est $cr_2 = 0,98$.

c) Pour les pompes E7-E8, ces coefficients sont :

$$D^{\circ}_3 = - 4570 ; \quad a_3 = 622 ; \quad b_3 = - 5 ;$$

Le coefficient correcteur est $cr_3 = 0,94$.

ii) Les éléments nécessaires au calcul de la h.m.t. :

Le calcul des hauteurs manométriques nécessite la connaissance des niveaux de l'ensemble des réservoirs (effacement et première élévation), mais surtout de la pression de refoulement à la sortie de l'usine.

f) La modélisation de la pression de refoulement :

i) La méthode de calcul approchée retenue :

Les résultats contenus dans le rapport LEQUENNE (juin 87) ont suggéré un modèle simplifié pour calculer la pression de refoulement. Ce modèle est un modèle statistique déterminé et calé à partir d'une recherche des facteurs explicatifs .

Le premier facteur explicatif est la pression statique du réseau prise comme la hauteur absolue du réservoir en équilibre sur ce réseau, c'est à dire celui de Montreuil, le réservoir de Gagny n'étant ouvert que quelques heures pendant la nuit.

Ce facteur est appelé HabsMON et a pour expression :

$$\text{HabsMON} = \text{Cote(radier)} - \text{cote(axe pompe)} + \text{niveau(R7,R8 ou R9)}.$$

Pour l'ancienne élevatoire, la différence des cotes est de 80,75 m et pour la nouvelle 83,7 m.

Le second facteur est la surpression "SURPRESS", qui représente la perte de charge due au débit fourni par l'usine au réseau, dont on admet qu'elle est modélisée de façon acceptable par la "surpression" moyenne de la configuration qui produit ce débit.

Cette surpression ne dépend que de la configuration et est égale à moyenne des sauts observés pour cette configuration lors de la remise en marche des pompes après l'arrêt des pointes EDF. Ces sauts de pression sont mesurés sur les rubans enregistrant en continu les pressions de refoulement des ancienne et nouvelle élévatoires (document en annexe II, du rapport sur PILOTE, mai 88).

Le dernier facteur est constitué par la consommation du réseau de première élévation. $\Delta P_{REF}(h)$ représente une correction fonction de l'heure correspondant aux mouvements réguliers d'eau dans le réseau que l'on ne peut, à ce stade de développement du projet que modéliser statistiquement:

$$\Delta P_{REF}(h) = 0,00022 * \text{tirage_horaire} - 2$$

La pression de refoulement $P_{REF}(h)$ à l'heure h peut être modélisée par :

$$P_{REF}(h) = \text{HabsMON} + \text{SURPRES} + \Delta P_{REF}(h)$$

d) Perte d'eau entre la nourricière et les réservoirs d'effacement

PILOTE considère que la perte moyenne d'eau, non comptabilisée dans les lavages de filtres, entre la nourricière et le réservoir d'effacement est de 6%.

e) L'évolution des réservoirs d'effacement :

i) L'aspiration par l'usine élévatoire :

L'ancienne élévatoire aspire l'eau dans les réservoirs R1 et R2. Le modèle considère que ces deux réservoirs battent ensemble. La nouvelle élévatoire aspire l'eau du réservoir R3.

ii) La répartition des apports de la nourricière :

Le modèle considère que les quantités d'eau en provenance de la nourricière seront réparties entre R1-R2 et R3 en proportion de ce qui prélevé par l'ancienne et la nouvelle élévatoire. Cette simplification est justifiée par la possibilité d'orienter aisément l'eau vers R1-R2 ou R3 par un jeu de vannes.

f) Evolution du niveau d'eau dans le réservoir de Montreuil :

Pendant les heures de fermeture de Gagny, le modèle considère que la variation du niveau de Montreuil est donnée par :

$$\Delta Nivmon = (VOLREF - DEMANDE) / SURFMON$$

VOLREF représente le volume total refoulé par l'élévatoire.

DEMANDE représente la consommation du réseau.

SURFMON représente la surface des réservoirs de Montreuil en service, c'est à dire habituellement SURF-R7 + SURF-R8 + SURF-R9.

2) L'aspect énergétique de la simulation :

a) Les consommations énergétiques des groupes élévatoires :

Comme pour le calcul des débits, le modèle considère que les consommations électriques des pompes d'une même famille sont identiques.

Comme seconde hypothèse, le modèle considère la fonction liant la puissance consommée au débit des pompes comme une relation linéaire du type :

$$PUISSANCE = a * DEBIT + b$$

Un certain nombre d'enregistrements simultanés de puissances et de débits ont été effectués pour l'ensemble des pompes. Les coefficients a et b de ces relations linéaires ont été déterminés à partir de ces enregistrements en recherchant la droite des moindres carrés. Le débit retenu dans l'équation est le débit calculé par l'automate, qui correspond à celui utilisé dans la simulation hydraulique de l'usine.

i) Les pompes E7-E8 :

i.a) La recherche des coefficients de la régression linéaire

Elle a été effectuée à partir d'une série d'enregistrements de la puissance active en fonction du débit calculé par l'automate. La droite des moindres carrés a pour équation :

$$P_{08} = D_{08} \cdot 583 \text{ Kwatts.s/m}^3 + 951 \text{ Kwatts.}$$

Le graphe donné en annexe III du rapport de mai 88, montre la courbe obtenue par la régression linéaire et deux autres droites :

- celle obtenue à partir de la courbe de pompe, d'équation :

$$P_{og} = D_{og}.412.Kwatts.s/m^3 + 1560 Kwatts.$$

- celle obtenue en prenant le débit mesuré à l'ultraflux :

$$P_{og} = D_{og}.599.Kwatts.s/m^3 + 1061 Kwatts.$$

ii) Les pompes E1-E6 et e11-e12 :

Pour ces deux familles, la recherche des coefficients a commencé par l'enregistrement, heure par heure, de la totalisation des compteurs d'impulsions et du débit calculé par automate pour chaque groupe en service. Il en résulte que :

ii.a) Pour les pompes E1-E6, l'équation linéaire est :

$$P_{o1} = D_{o1}.812.Kwatts.s/m^3 + 491 Kwatts.$$

ii.b) Pour les pompes e11-e12, l'équation linéaire est :

$$P_{o11} = D_{o11}.188.Kwatts.s/m^3 + 557 Kwatts.$$

b) La consommation énergétique de la nourricière :

Comme première approximation, PILOTE considère que la puissance appelée par la nourricière est proportionnelle à son débit, soit :

$$P_{nou} = D_{nou} * 183 kW.s/m^3$$

Le coefficient est obtenu à partir de la moyenne de la consommation énergétique de la nourricière sur le mois précédent.

c) Les ozoneurs :

PILOTE considère que la puissance appelée pour l'ozonation ne variera pas pour toute la durée de la plage horaire, sauf si cela est explicitement précisé. PILOTE gardera donc la valeur donnée initialement par les chefs de poste.

d) Le talon :

Le talon de consommation représente la somme des consommations énergétiques qui ne sont pas clairement identifiables ou calculables.

La puissance totale consommée par l'usine est enregistrée par pas de temps de 10 mn.

Le talon est alors obtenu en retirant de la puissance totale consommée par l'usine tout ce qui est calculable ou identifiable (élévatoire, nourricière, filtration, ozoneurs). Une première étude a montré que la forme de ce talon était à peu près identique d'une part pour les jours ouvrables, et d'autre part pour les jours de fin de semaine.

VI) LA MODELISATION DE LA PREVISION DE CONSOMMATION

A) Les modalités de la construction du modèle

Toute optimisation de gestion suppose une prévision et la recherche de stratégies de pompage n'échappe pas à cette règle. L'élaboration d'un modèle de prévision de consommation, servant de base à la gestion hydraulique de l'usine, a nécessité plusieurs phases :

- l'analyse des données brutes
- la construction de modèles à partir des traits caractéristiques apparus lors de l'analyse
- l'ajustement et la mise en oeuvre des modèles en situation réelle d'exploitation
- l'évaluation des modèles sur une période déterminée.

La phase d'analyse des données brutes correspond à l'étude faite par M. PENET, [PENET 87], et le présent document prend appui sur ses résultats.

Cette étude a conduit à la construction d'un modèle. Ce modèle commence par établir une prévision globale qui est ensuite décomposée. Il se base sur une régression linéaire pour obtenir la consommation journalière, cette valeur étant ensuite répartie heure par heure à l'aide de profils types. Ce modèle sera appelé plus loin dans le texte le modèle "PROFIL".

Une fois défini le principe de construction du modèle, ceux-ci ont été ajustés sur 3 périodes, en ce qui concerne les données de Neuilly/S/Marne.

a) Première étude correspondant à une période d'hiver.:

Période de calage : 1 décembre 1987 - 14 février 1988.

b) Seconde étude correspondant à une période de printemps

Période de calage : 1 mars 1987 - 31 mai 1987 + 1 mars 88 - 31 mars 88.

c) Troisième étude correspondant à une période d'été

Période de calage : 1 juin 87 - 31 juillet 87.

Les modèles ont été calés avec les données de l'année précédente.

B) L'analyse des données brutes

1) Les données utilisées

a) Choix de la variable dépendante :

La variable dépendante est le tirage défini comme la somme algébrique du volume d'eau refoulé par les élevatoires et des variations du volume d'eau contenu dans les deux réservoirs de 1ère élévation du réseau, c'est-à-dire Montreuil et Gagny. Les variations de volume des réservoirs de deuxième et troisième élévation ont été négligées.

b) Le choix du pas de temps :

Le pas de temps retenu est l'heure. D'autre part, on se servira également des bilans journaliers réalisés par le système de gestion des données de l'usine, SOPHIA. Dans SOPHIA, les tirages quotidiens sont égaux à la somme des tirages horaires de 6 heures du matin à 6 heures le lendemain matin.

c) Critiques sur les données :

Les données utilisées sont entachées d'incertitudes provenant pour la plupart d'un manque d'informations sur les mouvements d'eau d'un réseau de distribution vers les autres.

Les données sont également entachées d'un certain nombre d'erreurs de saisie, dont certaines ont pu être corrigées après un contrôle graphique de cohérence, réalisé à l'aide des fonctions graphiques du tableur LOTUS.

2) Les premières hypothèses formulées :

La méthodologie qui a servi de base à l'étude de PENET sur les consommations a reposé sur les hypothèses suivantes :

- les aspects sociaux sont prédominants et les variables météorologiques ont une influence complémentaire.

- les comportements sociaux sont homogènes selon les types de jour définis (pour des conditions météorologiques identiques, le tirage d'un vendredi devrait toujours être supérieur à celui du samedi, par exemple)

- les aspects sociaux seront d'abord pris en compte, et on fera apparaître l'influence des variables météorologiques sur un tirage "réduit", dégagé des aspects sociaux.

3) Des premières hypothèses à l'élaboration du modèle

L'analyse du rapport de M. PENET, sur les consommations d'eau pendant l'année 1986 a orienté la construction des modèles dans deux directions :

- une décomposition de l'année en 5 saisons
- une sélection de facteurs explicatifs pertinents

a) La décomposition de l'année en 5 saisons :

L'étude des profils quotidiens de consommation par l'analyse factorielle des correspondances et la nature des facteurs explicatifs ont conduit à décomposer l'année en 5 périodes, qui ne recoupent pas tout-à-fait les hypothèses initiales. Le mois de Juin a été rattaché à la période d'été, et l'automne a été séparé du printemps.

- hiver du 1er décembre au 28 février
- printemps du 1er mars au 31 mai
- été du 1er juin au 31 juillet
- août
- automne du 1er septembre au 30 novembre.

A chacune de ces périodes correspond des profils et des jeux de paramètres particuliers.

b) La sélection des facteurs explicatifs

Une série d'étude statistique a indiqué que la température maximale de la journée était le meilleur facteur météorologique explicatif. Toute introduction d'un autre facteur n'apportait que très peu de précisions au regard de la difficulté de gérer ce nouveau facteur.

Aussi, pour les 5 saisons, seule la température a été retenue comme facteur explicatif.

C) La construction des modèles profils

1) Le principe

L'établissement du modèle PROFIL s'opère en deux temps. Il faut d'abord caler les paramètres de l'équation linéaire donnant la consommation journalière, puis déterminer les profils types de consommation.

a) Recherche des paramètres de l'équation linéaire :

Le modèle de régression proposé est basé non sur les valeurs brutes des consommations, mais sur des valeurs réduites. Une consommation réduite pour un jour J est définie par le rapport de la consommation brute sur la moyenne des consommations des mêmes jours de la semaine, moyenne calculée sur la période considérée.

Par exemple, si l'on considère la période de calage du 1^{er} septembre au 15 novembre, la consommation réduite Cr du lundi 2 nov. sera :

$$Cr = \frac{\text{consommation brute du lundi 2 nov}}{\text{moyenne des consommations des lundis compris entre 1 sep et 15 nov.}}$$

La démarche à suivre pour calculer les coefficients de l'équation linéaire est donc la suivante :

- calcul des moyennes des consommations par jour de la semaine.
- calcul des consommations réduites.
- détermination des coefficients a, b et c de l'équation linéaire de régression. Cette détermination a été faite à l'aide du logiciel STATGRAF.

Le modèle commence donc par déterminer la consommation réduite prévisionnelle du jour par l'équation suivante, qui est appelée facteur-jour :

$$\text{facteur-jour}(\text{jourJ}) = a * \text{consoréduite}(\text{jourJ}-1) + b * T_{\text{prévue}} + c.$$

Il suffit ensuite de multiplier par la moyenne des consommations du jour pour avoir la prévision globale.

b) Recherche des profils types :

Un profil de consommation horaire est obtenu en divisant la consommation horaire par la consommation journalière. Pour chaque heure de la journée, correspond donc une fraction, exprimée en millièmes, de la consommation globale de la journée.

Avant de passer à l'analyse factorielle des correspondances, les consommations seront étudiées heure par heure pour chacun des jours.

Ensuite, on pourra aborder la phase d'analyse factorielle proprement dite pour opérer un regroupement de profils, et voir si ces résultats permettent une confirmation ou infirmation des hypothèses contenues dans le rapport PENET (juin 87).

c) Quelques points sur l'analyse factorielle des correspondances :

L'AFC est essentiellement un mode de présentation graphique de tableaux de contingence. Elle vise à rassembler en un ou plusieurs graphiques (généralement moins de 4) la plus grande partie possible de l'information contenue dans un tableau en s'attachant aux correspondances entre les caractères, c'est à dire aux valeurs relatives.

L'AFC fait partie des techniques qui réduisent l'information pour augmenter la signification d'un ensemble de données.

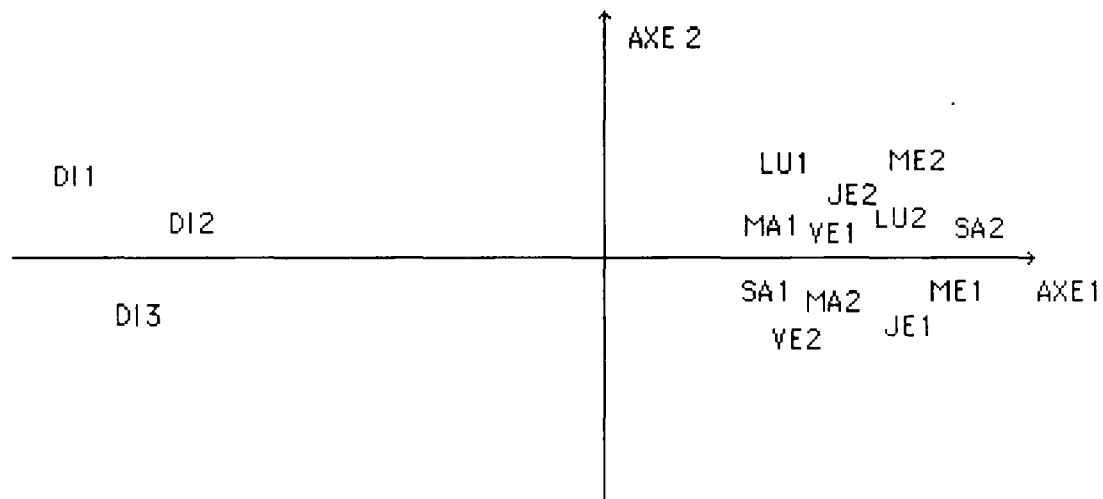
Le tableau de contingences est formé dans notre cas de N lignes représentant les N jours étudiés et de 24 colonnes représentant les consommations horaires pour chacun des N jours, codifiés par l'indication des deux premières lettres du jour de la semaine accolées au numéro de la semaine.

	*	jour	*	6 - 7 h	*	7 - 8 h	*	*	*	total	*
	*		*		*		*		*		*		*
1	*	LU1	*		*		*		*		*		*
2	*	MA1	*		*		*		*		*		*
3	*	ME1	*		*		*		*		*		*
4	*	JE1	*		*		*		*		*		*
...	*		*		*		*		*		*		*
N	*	DI10	*		*		*		*		*		*

d) Rappel des résultats principaux de l'AFC :

Une analyse factorielle des correspondances a été réalisée pour les saisons d'hiver, de printemps, d'été et d'automne 86.

Pour les quatre saisons, l'axe principal d'inertie révèle une opposition des profils du dimanche et des jours fériés à ceux des autres jours de la semaine.



Le second axe d'inertie met en opposition les profils du samedi à ceux des jours ouvrables.

Une fois que les dimanches, jours fériés et samedis ont été retirés de l'analyse, l'AFC ne révèle pas d'autre opposition de profils.

Aussi, les tous premiers essais de prévision ont été réalisés en se basant uniquement sur 3 profils correspondant aux dimanches + jours fériés, samedis et jours ouvrables. Mais les premières analyses des prévisions ont révélé des erreurs systématiques pour le lundi matin (sous-estimation) et la fin de nuit du vendredi au samedi (surestimation).

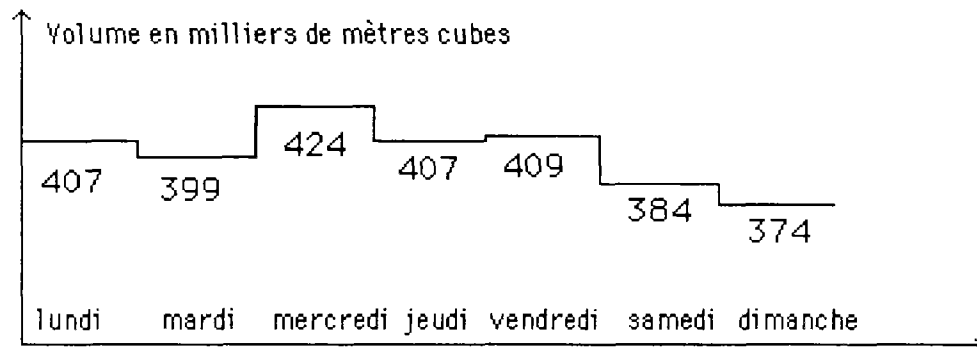
Cela conduisait à rajouter 2 profils particuliers (lundi et vendredi) et à ce niveau, il apparaissait plus simple de définir un profil par jour de semaine. C'est ce qui a été réalisé pour le printemps et l'été.

Le graphique suivant donne un exemple de profils pour les prévisions du printemps 88.

3) CARACTERISTIQUES DU MODELE PROFIL PAR SAISON

a) Pour l'hiver (du 1 décembre au 28 février):

a.1) Les consommations moyennes par jour de la semaine en m³

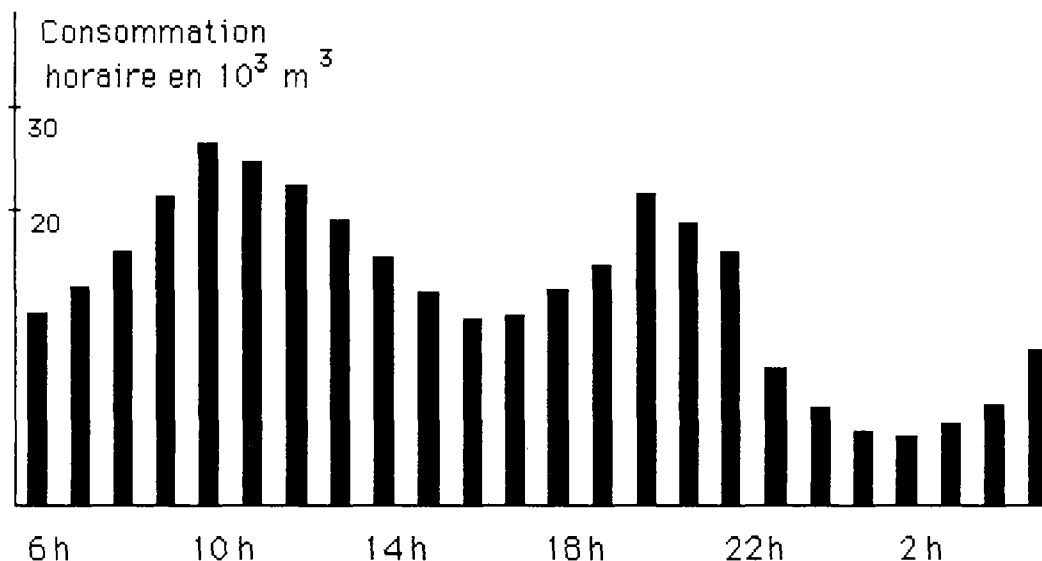


a.2) Les coefficients de l'équation de régression

$$\text{conso-réduite (J)} = a \times \text{conso-réduite (J-1)} + bT + c$$
$$a = 0,64, \quad b = 0,003, \quad c = 0,034$$

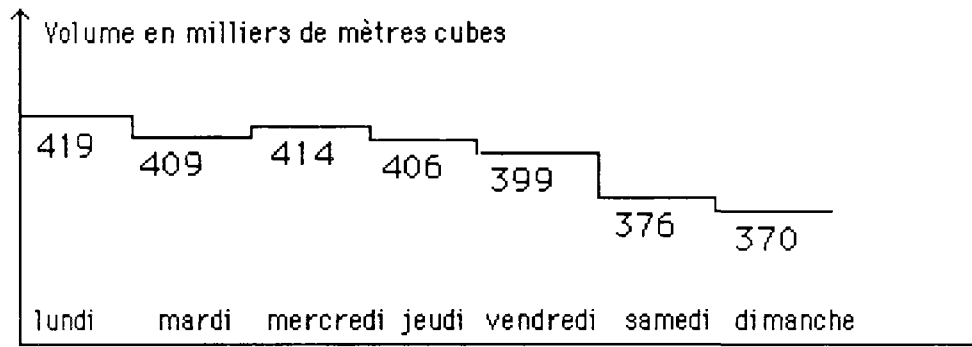
L'influence des vacances de Noël et de février a été prise en compte en introduisant le facteur multiplicatif empirique 0.9

a.3) Exemple de profils de consommation



b) Pour le printemps (du 1 mars au 31 mai) :

b.1) Les consommations moyennes par jour de la semaine en m³



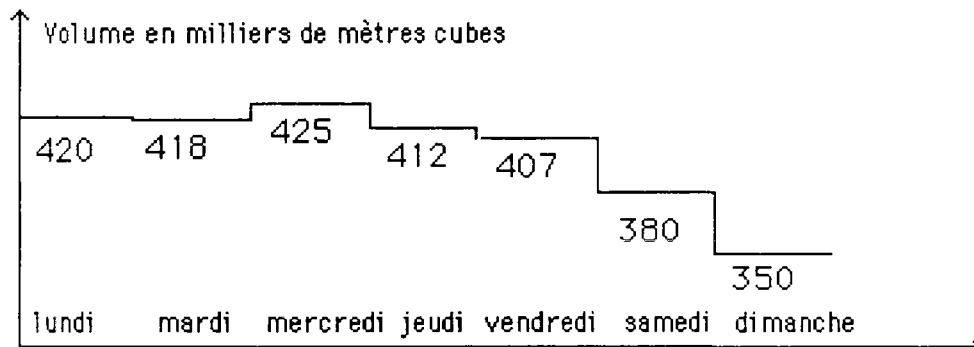
b.2) Les coefficients de l'équation de régression

$$\text{conso-réduite (J)} = a \times \text{conso-réduite (J-1)} + bT + c$$
$$a = 0,64, \quad b = 0,002, \quad c = 0,34$$

L'influence des vacances de Pâques a été prise en compte en introduisant le facteur multiplicatif 0.9

c) Pour l'été (du 1 juin au 31 juillet) :

c.1) Les consommations moyennes par jour de la semaine en m³

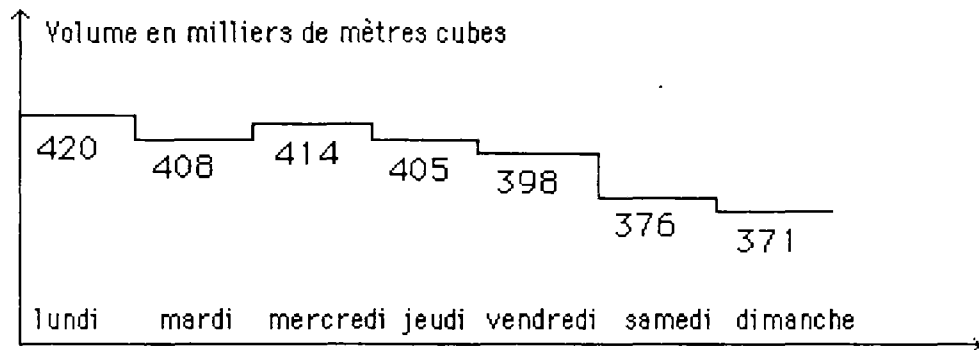


c.2) Les coefficients de l'équation de régression

$$\text{conso-réduite (J)} = a \times \text{conso-réduite (J-1)} + bT + c$$
$$a = 0,37, \quad b = 0,02, \quad c = 0,20$$

d) Pour l'automne (du 1 septembre au 30 novembre) :

d.1) Les moyennes par jour de la semaine en m³



d.2) Les coefficients de l'équation de régression

$$\text{conso-réduite (J)} = a \times \text{conso-réduite (J-1)} + bT + c$$
$$a = 0,47 \quad b = 0,004 \quad c = 0,47$$

L'influence des vacances de la Toussaint est prise en compte en introduisant le facteur multiplicatif 0,95

VII) LE CADRAGE DU SUJET ET DE SES LIMITES

A) L'ensemble des éléments constituant une stratégie de pompage

Une stratégie de pompage complète est composée des variables suivantes, à définir pour chaque pas de temps :

- configuration de l'usine élévatoire,
- indication de régime nourricier,
- coefficient de répartition de l'eau entre l'ancienne et nouvelle élévatoire
- indication sur le planning de lavage des filtres,
- indications sur les purges.

Pour définir complètement une stratégie, il faut encore indiquer le pas de temps avec lequel on va travailler.

B) Le sujet et ses limites

Compte-tenu du contexte de développement, (élève-chercheur seul), et de l'objectif de la thèse, qui était d'illustrer une méthodologie, il a été décidé de ne prendre en compte que 2 variables de commandes, la configuration élévatoire, et l'indication de régime nourricier.

Le pas de temps horaire a été retenu parce qu'il correspond à celui avec lequel la très grande majorité des informations sont saisies sur l'usine, comme par exemple la totalité des niveaux des réservoirs. D'autre part, ce choix de l'heure n'est pas pénalisant au niveau de la finesse des décisions, car le nombre de changements de configurations dépasse rarement 5 ou 6 par jour.

Ch2 L'APPORT DE L'UTILISATION DE CONNAISSANCES DANS LES TECHNIQUES DE RESOLUTION DE PROBLEMES

I) L'INTRODUCTION DE LA NOTION D'HEURISTIQUE

A) Les techniques d'exploration systématique de l'espace

1) Les espaces d'états

En intelligence artificielle, la notion de résolution de problèmes est étroitement associée aux techniques de recherche dans les graphes [FARRENY, 87]. Ce type de problème à résoudre est formé :

- d'un état initial,
- d'un état final, (ou d'un ensemble d'états terminaux)
- d'un ensemble d'opérateurs permettant de passer d'un état à un autre.

Un état lui-même se définit comme une description de la situation initiale dans laquelle le problème est posé, ou d'une situation intermédiaire quelconque. Cette description correspond à l'ensemble des données qui interviennent dans le problème, et qui peuvent être structurées sous différentes formes.

La résolution de ces problèmes passe par l'exploration d'un espace de représentation, et la définition d'une représentation est donc une opération capitale. Une représentation est définie comme un ensemble de conventions sur la façon de définir les objets, leurs regroupements, et leurs relations.

2) Exemples de problèmes définis dans des espaces d'états

Un certain nombre de problèmes types sont régulièrement mentionnés. Parmi ceux-ci on peut citer :

- le problème des 8 reines, qui consiste à trouver une configuration de ces reines sans prise sur un échiquier,
- le problème des tours de Hanoï, où l'on dispose de 3 axes verticaux. Sur le premier sont disposés N disques de rayon décroissant, et le problème consiste à reconstituer l'empilement sur le 3^{ème} axe, en déplaçant les disques un par un, sans jamais superposer un disque sur un plus petit.
- le problème du voyageur de commerce, qui consiste à définir le parcours minimal entre N villes où le voyageur doit passer une et une seule fois,

- le jeu du taquin qui est détaillé plus loin.

Plusieurs auteurs, dont FARRENY, ont mis en évidence des caractéristiques générales de ce type de problèmes, et des constantes nécessaires à leur résolution :

- un ensemble de moyens de transformation du problème, c'est à dire d'opérateurs permettant de passer d'un état à un autre,
- un moyen de tester les situations obtenues par rapport à l'objectif.

La représentation d'un problème consiste à définir formellement ces constantes, c'est à dire à préciser dans quel formalisme elles seront décrites. Une bonne représentation est celle qui rend explicite les faits importants et expose clairement les contraintes naturelles inhérentes au problème. Elle doit être complète, concise, transparente et exécutable.

4) L'exemple du jeu du taquin

Ce jeu se déroule sur un damier carré de 9 cases, comprenant 8 pions numérotés de 1 à 8 et une case vide.

Dans ce jeu, la bonne représentation consiste à définir les opérateurs pour qu'ils servent à déplacer la case vide, vers le haut, la bas, la gauche ou la droite. Suivant que sa position est le centre, le milieu d'un côté, ou un coin, il y a 4, 3, 2 mouvements possibles. Le formalisme, consistant à gérer le déplacement des 8 cases, serait bien sûr beaucoup trop lourd.

Le problème du jeu consiste à définir une séquence optimale de déplacements dans l'espace d'état, qui permette de passer de la situation initiale à une situation finale.

L'exemple qui sera mentionné plus loin est caractérisé par l'état initial et le but suivant :

```
*----*----*----*
* 2  * 4  * 3  *
*----*----*----*
* 1  *      * 5  *
*----*----*----*
* 6  * 7  * 8  *
*----*----*----*
* situation initiale
```

```
*----*----*----*
* 1  * 2  * 3  *
*----*----*----*
* 4  *      * 5  *
*----*----*----*
* 6  * 7  * 8  *
*----*----*----*
*          but      *
```

A partir de cette situation initiale, 4 opérateurs sont utilisables :

- déplacer la case vide vers le haut :

Configuration C1

```
*----*----*----*
* 2 *      * 3 *
*----*----*----*
* 1 * 4 * 5 *
*----*----*----*
* 6 * 7 * 8 *
*----*----*----*
```

- déplacer la case vide vers le bas :

Configuration C2

```
*----*----*----*
* 2 * 4 * 3 *
*----*----*----*
* 1 * 7 * 5 *
*----*----*----*
* 6 *      * 8 *
*----*----*----*
```

- déplacer la case vide vers la gauche

Configuration C3

```
*----*----*----*
* 2 * 4 * 3 *
*----*----*----*
*      * 1 * 5 *
*----*----*----*
* 6 * 7 * 8 *
*----*----*----*
```

- déplacer la case vide vers la droite

Configuration C4

```
*----*----*----*
* 2 * 4 * 3 *
*----*----*----*
* 1 * 5 *      *
*----*----*----*
* 6 * 7 * 8 *
*----*----*----*
```

Une recherche complète du meilleur chemin donne pour chacune des 4 configurations les résultats suivants :

- C1 : but atteint au mieux en 3 coups.
- C2 : but atteint au mieux en 5 coups.
- C3 : but atteint au mieux en 5 coups.
- C4 : but atteint au mieux en 5 coups.

B) Recherches aveugles dans les espaces d'état

1) Le formalisme habituel

S'il y a plus d'une possibilité d'application d'un opérateur, il devient indispensable de définir une stratégie d'exploration de l'espace de recherche.

On peut reprendre le formalisme habituel avec les 3 termes, ARBRE, NOEUD OUVERT et NOEUD FERME.

Un noeud ouvert est un noeud dont la prolongation vers un état suivant est possible, ce qui n'est pas le cas d'un noeud fermé.

La situation initiale est donnée par :

ARBRE = état initial,
NOEUDS OUVERTS = état initial,
NOEUDS FERMES = ensemble vide.

La boucle de raisonnement est formée des étapes suivantes :

Si NOEUDS OUVERTS = Ensemble vide, Echec.

Sinon, choisir un noeud N dans NOEUDS OUVERTS, et le placer dans NOEUDS FERMES,

Si le développement de N est le but, Succès,

Sinon, placer les successeurs de N dans NOEUDS OUVERTS.

Il est évident qu'à chaque étape du raisonnement, il peut y avoir un choix à effectuer. Ce choix peut être automatisé en utilisant un algorithme de recherche. L'état choisi peut être complètement développé, c'est à dire que tous les successeurs sont engendrés et toutes les alternatives sont considérées pour participer aux choix suivants. Mais, le plus souvent, on s'arrête au premier successeur engendré, on note pour un retour éventuel que cette étape n'a pas été entièrement explorée, et les nouveaux états s'ajoutent aux alternatives connues.

Parmi ces algorithmes d'exploration systématique, la recherche en largeur d'abord, et la recherche en profondeur d'abord, sont les plus utilisés.

2) Recherche en largeur d'abord

Dans cet algorithme, à chaque étape de l'exploration, c'est le noeud le plus ancien placé dans l'ensemble des noeuds à ouvrir qui est développé.

Dans l'exemple précédent du jeu du taquin, les deux premières étapes conduiraient aux situations suivantes.

L'étape 1 consiste à ouvrir le premier noeud du premier niveau. Il en résulte la situation suivante :

ARBRE = état initial,
NOEUDS OUVERTS = C2, C3, C4, successeurs de C1
NOEUDS FERMES = C1.

L'étape 2 consiste à ouvrir le second noeud du premier niveau. Il en résulte la situation suivante :

ARBRE = état initial,
NOEUDS OUVERTS = C3, C4, successeurs de C1, successeurs de C2
NOEUDS FERMES = C1, C2.

Cet exemple illustre que dans cette stratégie, on ne descend d'un niveau dans l'arbre que lorsque tous les noeuds du niveau précédent ont été développés.

3) Recherche en profondeur d'abord

Dans cet algorithme, à chaque étape de l'exploration, c'est le noeud le plus récent placé dans l'ensemble des noeuds à ouvrir qui est développé.

Dans l'exemple précédent, les deux premières étapes conduiraient aux situations suivantes.

La première étape est identique au cas précédent. L'étape 2 consiste à ouvrir le premier noeud du second niveau. Il en résulte la situation suivante :

ARBRE = état initial,
NOEUDS OUVERTS = C2, C3, C4, successeurs de C1, successeurs des successeurs de C1,
NOEUDS FERMES = C1, successeurs de C1.

Cet exemple illustre que dans cette stratégie, on descend d'un niveau dans l'arbre à chaque exploration.

C) Limites des recherches aveugles et introduction des heuristiques

Les algorithmes très simples qui explorent systématiquement l'arbre de représentation du problème nécessitent un temps de calcul proportionnel au

nombre de noeuds générés. La complexité des algorithmes est de l'ordre de l^p , l étant le nombre moyen de descendants par noeud, et p la profondeur de la solution. Dans le cas de la configuration donnée au jeu du taquin, où l'on a en moyenne 3 possibilités par coup, une recherche aveugle de la solution nécessiterait d'envisager 3^4 coups.

Ces algorithmes d'exploration systématique ne prennent pas en compte les informations sur une situation donnée à un niveau du raisonnement. Le choix d'un noeud à développer est comparable à un processus de prise de décision, et, en temps que tel, nécessite de disposer des informations qui permettent d'orienter un choix. Ces informations sont dans certains cas immédiatement disponibles, mais souvent, il faut les créer. .

C'est le rôle des heuristiques de savoir utiliser et au besoin créer des informations pour orienter l'exploration de l'arbre de représentation du problème. Les heuristiques ont deux manières principales d'orienter les choix. La première consiste à définir le noeud à développer, et la seconde à indiquer quels noeuds il est possible ou indispensable d'éliminer.

La génération d'informations peut être réalisée par différents moyens. Le premier consiste à utiliser une fonction d'évaluation. Une telle fonction calcule une estimation de la distance entre un état considéré et le but, généralement sous la forme :

$$Et(\text{longueur totale}) = d(\text{déjà parcouru}) + Es(\text{distance restante}).$$

Toute estimation doit être obligatoirement minorante. La raison est que, si l'on trouve un chemin complet, en développant toujours le chemin de plus faible sous-estimation, la recherche peut s'arrêter lorsque toutes les estimations de chemins incomplets sont supérieures à celle d'un chemin complet.

D) Exemples de fonctions d'évaluation

1) Exemple du jeu du taquin

Dans ce jeu, on peut citer plusieurs fonctions d'évaluation.

a) La comptabilisation des cases mal placées

La première fonction consiste à indiquer le nombre total N de cases avec chiffre, qui ne sont pas à la place qu'ils doivent occuper dans la configuration finale. Dans l'exemple donné, pour la situation initiale, ce nombre $N = 3$.

b) Le calcul de la distance de Manhattan

Une seconde fonction comptabilise la somme S minimale totale des déplacements à effectuer pour amener chaque case dans sa position finale. Les déplacements ne peuvent avoir lieu que sur des axes perpendiculaires. Cette fonction est appelée le calcul de la distance de Manhattan, et dans l'exemple considéré, la somme $S = 1+1+0+2+0+0+0+0+0=4$.

c) Le calcul de l'écart à une séquence type

Une troisième fonction consiste à mesurer l'écart à une séquence type définie comme la configuration but, après une ou plusieurs rotations élémentaires autour de la case centrale vide. C'est à dire, que, quelles que soient leurs positions, la case 1 doit suivre la 4, la 2 la 1, la 3 la 2, etc... La fonction compte 2 points pour chaque pion non suivi par le "bon", et 1 si la case vide n'est pas au centre. Dans notre cas, cette séquence $Se = 6$.

d) Le calcul d'une pondération de fonctions d'évaluation

Une quatrième fonction d'évaluation consiste à calculer une somme pondérée des 2^{ème} et 3^{ème} fonctions d'évaluation. Des auteurs [DORAN et MICHIE, 69], se sont penchés sur la pondération, et ont retenu les valeurs 1 et 3 pour la pondération, soit $F = S + 3*SE = 22$.

2) Remarques sur les fonctions d'évaluation

Les exemples précédents montrent clairement qu'il n'y a pas une seule et unique fonction d'évaluation par type de problème, mais que l'on peut en définir un très grand nombre. La finalité étant bien sûr d'utiliser la fonction la mieux adaptée au problème.

3) Comparaison des décisions prises sur l'exemple donné

En prenant l'exemple précédent, on aboutit au tableau décisionnel suivant, pour choisir l'opérateur à appliquer dans la situation initiale.

	*	C1	*	C2	*	C3	*	C4	*
	*	-----	*	-----	*	-----	*	-----	*
N	*	4	*	5	*	4	*	5	*
	*	-----	*	-----	*	-----	*	-----	*
S	*	3	*	5	*	4	*	5	*
	*	-----	*	-----	*	-----	*	-----	*
Se	*	3	*	9	*	7	*	9	*
	*	-----	*	-----	*	-----	*	-----	*
F	*	12	*	32	*	25	*	32	*
	*	-----	*	-----	*	-----	*	-----	*

Si l'on se servait de ces fonctions d'évaluation pour envisager de développer un noeud, on voit sur cet exemple que les fonctions S, Se et F donnent la meilleure solution (C1) alors que le calcul de N ne permet pas de départager entre les solutions C1 et C3.

Mais le remplissage de ce tableau se fait par des calculs plus ou moins simples. La relative complexité du calcul de la fonction d'évaluation F fait comprendre que, si le premier critère dans la construction d'une fonction est sa pertinence, un second critère est constitué par sa simplicité.

D) Complexité et efficacité des fonctions d'évaluation

La première fonction N donnée en exemple est simple, puisqu'elle consiste seulement à balayer 9 cases, en effectuant un seul test par case.

La seconde fonction S demande un calcul plus complexe, puisqu'il faut commencer par comptabiliser 9 distances élémentaires, en recherchant à chaque fois l'écart entre les positions initiale et finale, tout déplacement en diagonale étant interdit. Puis ensuite, il faut additionner ces déplacements.

La troisième fonction Se impose davantage de calculs et de tests, puisqu'il est nécessaire de commencer par des tests de positionnement de la séquence type avant de faire une comptabilisation des écarts.

Il est évident que la quatrième fonction F est plus complexe, puisqu'il faut avoir précédemment calculé les sommes S et Se.

La complexité des fonctions d'évaluation est d'ailleurs liée au nombre de contraintes non relaxées. Dans le cas du calcul du nombre N, tout se passe comme si toute case pouvait être déplacée en une seule fois de sa position d'origine à son but, ce qui correspond à une relaxation importante des contraintes, et donc à une sous-estimation des distances à parcourir.

Dans le cas de la seconde fonction, tout se passe comme si les déplacements des cases, limités à deux axes perpendiculaires, pouvaient se chevaucher. Ceci correspond à une relaxation des contraintes moindre que celle du cas précédent, mais correspond toujours à une sous-estimation de la distance. Pour la fonction Se, on suppose que, une fois la bonne séquence construite, la rotation à la configuration but est immédiate.

L'exemple de ces fonctions d'évaluation citées dans le jeu du taquin met en évidence que toute recherche d'informations, au travers d'un processus de prise de décision, a un coût. Si l'on approfondit la réflexion, on s'aperçoit qu'il

devient indispensable de mettre en relation l'effort d'acquisition de cette information, en prenant en compte la complexité de la fonction d'évaluation, avec son efficacité.

Une bonne exploration de l'espace de recherche consiste à utiliser la fonction d'évaluation adaptée à chaque niveau de l'exploration de l'espace de recherche. Quand le nombre d'alternatives possibles est important, les fonctions les plus précises ne sont pas obligatoirement les plus adaptées, car elles imposent des calculs lourds à des solutions qui auraient pu être rejetées par des fonctions plus simples. Il peut d'ailleurs être judicieux d'élaguer l'arbre de représentation du problème, comme cela est réalisé dans l'algorithme DENDRAL présenté dans les paragraphes suivants.

II) RESOLUTION DE PROBLEMES EN UTILISANT DES HEURISTIQUES POUR RESTREINDRE L'ESPACE DE RECHERCHE

A) L'élagage d'un arbre de recherche

Dans le cas précédent, il est montré comment, à l'aide d'une fonction d'évaluation qui crée des informations, des heuristiques peuvent contribuer à la sélection du meilleur noeud à développer, à chaque étape de résolution du problème.

Une autre possibilité de restreindre l'espace de recherche est de couper certaines branches de l'arbre généré, toujours en fonction des informations disponibles à une étape du raisonnement.

Le programme DENDRAL, [BUCHANAN, FEIGENBAUM 78] décrit dans le paragraphe suivant, est un exemple de méthodologie de résolution de problèmes, élagant un arbre de recherche qui a été généré à l'étape précédente du raisonnement. Ce programme génère des ensembles de solutions qu'il teste ensuite.

B) Description du programme DENDRAL de recherche d'une formule chimique

1) Présentation générale de DENDRAL

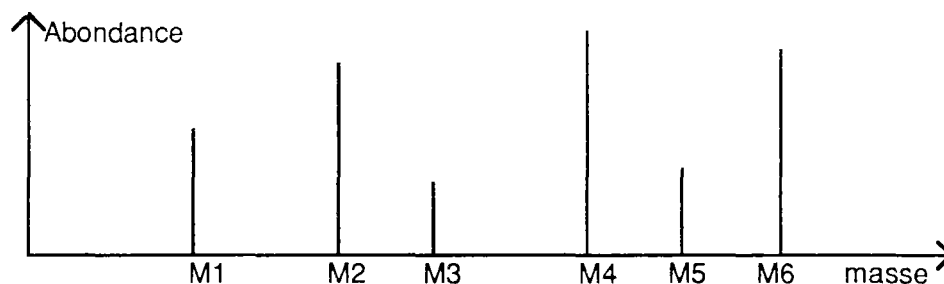
Le programme recherche les formules tridimensionnelles des molécules dont il connaît la composition chimique et les résultats d'une spectrographie de masse. Il a été développé par Buchanan à l'université de Stanford, et constitue la référence dans les programmes d'apprentissage par génération et test d'hypothèses.

Le système part de la connaissance d'une formule brute et d'un spectrographe de masse, et le but est de déterminer la formule développée du composé. Les informations initiales sont relativement nombreuses, mais leur interprétation est difficile.

a/ La spectrographie de masse

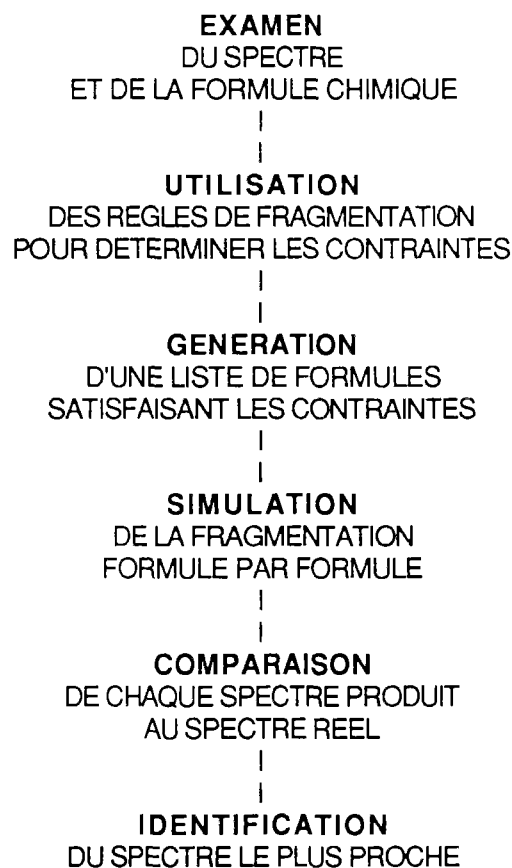
Un spectromètre de masse bombarde des molécules avec des électrons de haute énergie. Ces électrons cassent les molécules en fragments de masse plus petites.

Le spectromètre mesure alors l'abondance A relative des différents fragments de masse M déterminée, et présente l'information sous la forme d'un graphe dont un exemple est donné ci-dessous.



La façon dont une molécule se fragmente dépend de sa structure, aussi l'analyse du spectre va-t-elle donner des informations interprétables dans la recherche de sa formule initiale. Le problème est que l'examen et l'étude d'un spectre ne permettent pratiquement jamais de déduire sans ambiguïté une et une seule formule développée. Aussi les informations initiales vont elles servir de base à un traitement dont le schéma est décrit dans le paragraphe suivant.

b/ Les étapes d'une session de DENDRAL (recherche de formule)



2) Les étapes de résolution du problème

a) La planification par examen du spectre

Elle signifie, dans ce contexte, redéfinir le problème en termes qui réduisent l'effort de résolution du problème.

Le problème de déterminer toutes les combinaisons possibles d'un ensemble d'atomes est réduit à celui de la recherche de toutes les combinaisons cohérentes avec les contraintes définies par la spectrométrie de masse, en tant que technique. Cela signifie que certaines ruptures de liaisons entre atomes sont invraisemblables, comme les ruptures Carbone-Hydrogène par exemple.

L'inférence automatique de ces contraintes est la partie de planification des heuristiques de DENDRAL.

b) L'utilisation des règles de fragmentation

Les informations issues de l'examen définissent deux autres séries de contraintes, liées aux possibilités de groupements des atomes :

- les groupements d'atomes qui doivent faire partie de la formule développée,
- les groupements d'atomes qui ne peuvent en faire partie.

La planification est réalisée à l'aide d'un certain nombre de règles d'inférence interprétant les caractéristiques des spectres, par exemple :

$$\text{Si } x_1 + x_2 = M + 28$$

Alors la molécule doit contenir un groupement Cétone.

Le jeu de règles crée deux listes. La liste des composés interdits provient à la fois de l'analyse du spectrogramme et de connaissances générales sur les configurations instables.

Les heuristiques utilisent donc des informations directement issues de l'analyse du cas étudié, mais aussi des lois générales sur la chimie moléculaire, tout à fait indépendantes du spectrogramme étudié. Ce premier traitement des informations initiales est insuffisant pour définir une structure unique. Aussi un programme va-t-il créer des informations pour continuer le raisonnement, en générant des structures cohérentes avec les informations initiales, et en simulant leur fractionnement dans un spectromètre de masse.

c) La génération des structures possibles

Sous les contraintes de planification, l'algorithme de DENDRAL génère seulement les formules ou structures qui n'ont pas de parties interdites.

CONGEN (le programme générateur de contraintes), était d'abord implémenté comme un système autonome, et les contraintes étaient fournies par les chimistes utilisant le programme.

Le premier but de CONGEN était de permettre à l'utilisateur de spécifier de façon interactive certains types d'informations structurales obtenues à partir de l'une des sources différentes comme la spectroscopie, la dégradation chimique, etc...

Le second but est de générer une liste de structures, exhaustive et non redondante, consistante avec cette information.

De façon schématique, les algorithmes de CONGEN peuvent être vus comme réalisant une réduction de problèmes et une reconstruction ou une recomposition de sous-problèmes sur les structures moléculaires.

Les types de contraintes sont les symétries, les valences des atomes, les heuristiques sur les compatibilités entre groupements.

L'expérience montre qu'un bon générateur doit présenter plusieurs caractéristiques :

- la complétude,
- la non-redondance,
- l'utilisation optimale des contraintes connues.

CONGEN fonctionne en construisant par étape les structures candidates en ajoutant les groupes un par un.

d) La simulation

Cette partie du programme génère un grand nombre d'informations pour toutes les structures candidates en simulant leur comportement dans un spectromètre de masse. Ces informations vont servir de base de comparaison avec le spectromètre initial.

e) La comparaison

Les structures résultantes sont rangées dans l'ordre de ressemblance de leur spectre avec le spectre obtenu expérimentalement.

MSPRUNE, le programme d'élimination et de classement gère une liste de structures candidates et le spectre de masse de la molécule inconnue.

Il utilise une théorie simplifiée de la spectrométrie de masse pour prédire les fragmentations normalement attendues dans chaque structure candidate. Les spectres qui dévient le plus du spectre expérimental sont éliminés.

MSRANK utilise des règles plus élaborées pour classer les structures restantes, selon le nombre des pics prévus effectivement trouvés (ou non trouvés) dans le spectre expérimental. Ce classement est pondéré par l'importance des pics considérés.

IV) L'INTERET D'UTILISER DES HEURISTIQUES DANS LE CAS DU PROBLEME POSE

A) La notion d'heuristique

Après la phase de choix et de formulation d'une représentation d'un problème, on aborde celle de sa résolution proprement dite. Dans tous les cas, l'étape de résolution repose sur deux éléments clefs :

- le recensement des alternatives possibles,
- le choix de l'une ou l'élimination d'une partie d'entre-elles.

La notion d'heuristique intervient au niveau du choix et comprend l'ensemble des éléments qui améliorent l'efficacité de ce choix, par une exploration rationnelle de l'espace de recherche, en évitant les recherches systématiques coûteuses.

L'utilisation de fonctions d'évaluation ou l'algorithme "générer et tester" sont deux exemples qui indiquent comment des heuristiques peuvent structurer et guider un processus de recherche dans un espace.

Dans les cas décrits précédemment, les heuristiques sont constituées à la fois par la connaissance d'une ou plusieurs fonctions d'évaluation, et celle de leur utilisation, dans les cas où plusieurs fonctions sont envisageables.

B) La provenance des heuristiques

La provenance des heuristiques est diverse. Elles peuvent avoir pour origine une expérience accumulée, ou une compréhension synthétique du problème. Dans le cas du jeu du taquin, la fonction *Se* vient d'une compréhension de l'importance de la notion de séquence, pour la résolution du problème. La pondération *F* résulte par contre de toute une série de tests, sur un très grand nombre d'exemples.

Dans le cas du logiciel DENDRAL, c'est l'expérience accumulée par les chimistes qui a servi à définir les règles d'élagage et les modèles de génération d'informations.

Les heuristiques ne sont pratiquement jamais immédiatement accessibles et explicites, et leur formulation demande, dans pratiquement tous les cas, un long travail d'extraction et de mise en forme.

C) L'intérêt d'utiliser des heuristiques pour la planification

La détermination de stratégies de conduite hydraulique, même limitée aux configurations des usines élévatoire et nourricière, sur un pas de temps

horaire, et pour une durée de 24 heures, nécessite l'exploration d'un arbre de représentation très important. Ainsi, pour la seule usine élévatoire, 6324 configurations de pompage différentes sont possibles.

Aussi, utiliser une approche basée sur l'expérience de conduite a pour objet d'explorer plus rationnellement un espace de représentation de cette taille.

D'autre part, les méthodes de recherche opérationnelle, utilisent des fonctions de coût, intégrant des éléments aussi différents qu'une facture EDF, et qu'un risque de rupture d'approvisionnement en eau.

Un second objectif de la thèse était de rechercher des heuristiques concernant la planification, et de les intégrer dans une méthodologie qui soit à la fois capable de focaliser l'exploration et de rendre explicite la démarche de choix des stratégies.

Une part importante du travail de thèse a donc consisté à recueillir et formaliser l'expérience accumulée sur l'usine. C'est cette phase qui est décrite dans le chapitre suivant.

CHAPITRE 3

LE RECUEIL DE L'EXPERTISE

I) CARACTERISTIQUES DES SCHEMAS DE RAISONNEMENT D'EXPERTS

A) Quelques éléments sur l'ingénierie de la connaissance

Tout projet en intelligence artificielle qui a pour but de réaliser une tâche ou une activité donnée doit reposer sur des modèles de représentation de la connaissance. La cognitique est la science qui a précisément pour objet de cerner, extraire, et organiser l'ensemble des connaissances permettant de réaliser cette activité.

Toute activité fait appel à des connaissances de nature diverse, telles que les classifications, les schémas d'interprétation,... et qui sont contenues dans des endroits aussi variés que des écrits, des procédures, des schémas d'organisation ou encore dans la matière grise de certains individus. Il en résulte que l'expertise n'est pas uniforme, parce que les expériences qui ont servi à la construire ne sont pas uniformes, et encore moins la perception et l'intégration de ces expériences par les différents individus. Tout transfert d'expertise doit donc intégrer un aspect uniforme. Il n'en demeure pas moins que certaines lignes directrices apparaissent dans les stratégies de recueil.

J. P. COT, dans le Monde informatique du 22 septembre 86, distingue trois grandes catégories de connaissances.

La première est constituée par la connaissance publique, qui réside essentiellement sous forme de documents écrits (manuels, articles, schémas,...).

La seconde est la connaissance partagée, "qui permet à un groupe de personnes travaillant ensemble de se comprendre, sans avoir tout à se dire ou à s'expliquer les uns aux autres). Des activités mettant en jeu des connaissances partagées font appel à des interprétations communes du contexte. Il apparaît ici, une première fois, l'idée de contexte que l'on retrouvera très souvent par la suite.

La troisième catégorie est constituée par la connaissance privée, liée à l'individu, qui intègre des perceptions personnelles. Celles lui permettent d'extrapoler, à partir de son expérience, les procédures à suivre à partir du petit nombre d'éléments qu'il sait vite identifier.

Dans ce même article, J. P. COT indique aussi une autre typologie de la connaissance qui peut s'exprimer sous forme déclarative, procédurale, mais aussi "spatiale, visuelle, kinesthétique (sensation de mouvement éprouvés par le corps humain)".

Ainsi, par ses objectifs et ses domaines d'application, la cognitique se situe au croisement des sciences sociales et humaines, c'est à dire entre la sociologie, l'ethnologie, la systémique.

La référence à la sociologie vient du fait que toute expertise est inscrite dans un environnement, ne serait-ce que le lieu et les conditions de travail auxquels tout expert est soumis.

Décrypter toute une série de processus intériorisés dans les pratiques de l'expert s'apparente aux travaux des ethnologues qui cherchent à pénétrer, sans perturber, dans des systèmes de références, de codes, très éloignés des cultures standard. Ceci explique d'ailleurs la présence, parmi les cogniticiens reconnus, de plusieurs personnes qui ont une formation et une pratique de l'ethnologie. C'est le cas de VOGEL qui est cité plus loin.

La première question à laquelle doit répondre l'ingénierie de la connaissance est comment acquérir la connaissance experte et la représenter dans des formes manipulables par un ordinateur.

Ensuite, cette ingénierie doit savoir comment extraire, articuler et implémenter cette connaissance dans un système opérationnel.

B) Les acteurs d'un domaine de connaissance

1) Les approches d'un expert et d'un spécialiste

La connaissance d'un domaine particulier est rarement détenue par une seule personne, et elle est souvent liée à l'utilisation qui en est faite. Ainsi, sur un même domaine, on retrouve des experts et des spécialistes, qui ont des perceptions et des approches différentes de la connaissance.

L'opposition entre l'expert et le spécialiste d'un même domaine recouvre le plus souvent une opposition entre un modèle théorique et un modèle pratique de compétence. Schématiquement, le spécialiste cherche à affiner ou augmenter la connaissance d'un domaine, alors que la préoccupation principale d'un expert est d'utiliser cette connaissance.

L'expert [VOGEL 88] met en oeuvre le modèle le plus économique, le plus pratique et le plus opérationnel dans la majorité des cas qu'il doit traiter.

L'expert [WATERMAN 85] d'un domaine donné est capable d'extraire les notions clefs d'ensemble d'informations apparemment sans rapport, et sait aussi reconnaître un problème comme une instance d'un type de problème déjà connu, sur lequel il a donc déjà préalablement réfléchi.

2) Le rôle du cogniticien

Le premier rôle d'un cogniticien est d'identifier les caractéristiques du problème posé, et définir si sa résolution fait appel principalement à de l'analyse, à de la synthèse, ou aux deux. Par exemple, les problèmes de classification, de diagnostics, d'interprétation font appel à l'analyse, alors que la configuration, le "design", la planification demandent une approche plus synthétique.

Un second rôle du cogniticien est d'identifier, de contrôler et de clarifier les différents niveaux de perception et de construction du modèle. Il doit en effet être capable de rendre explicite la démarche suivie par l'expert pour produire ses résultats.

C) Points communs des raisonnements d'experts

Toute expertise est un découpage du réel, et représente une connaissance dont la caractéristique principale n'est pas d'être exhaustive mais opérationnelle. Plusieurs auteurs ont étudié le comportement d'un expert en essayant de décomposer son activité.

1) Les travaux précédents sur une démarche experte

a) Le principe de la rationalité limitée

D'après leur principe de la rationalité limitée [NEWELL et SIMON 1972], la capacité humaine à raisonner est limitée du fait de son inaptitude à traiter un grand nombre d'informations à la fois. Pour travailler efficacement avec toute la connaissance disponible sur un sujet, avec toutes les données recueillies, un cerveau humain est amené, d'une part à sélectionner les données les plus pertinentes, et d'autre part à schématiser par des représentations simples un certain nombre de situations types.

b) Les travaux sur la réalisation d'un diagnostic [PAUKER, KASSIRER 76]

D'après PAUKER et KASSIRER, l'activité d'un expert se décompose en deux phases :

- acquisition d'un nombre d'informations minimales, par prise en compte des faits ou combinaisons de faits saillants.

- génération d'un premier noyau d'hypothèses actives, généralement formé de 4 ou 5 hypothèses (rarement plus de 7) qui structurent la recherche ultérieure d'informations complémentaires.

c) Les travaux sur les jeux d'échec

Les expériences sur les joueurs d'échecs [DE GROOT 65], montrent que si l'on présente à un débutant et à un maître une configuration de milieu de partie, pendant un temps très limité (5 secondes), le maître peut reconstituer cette configuration, alors que le débutant ne peut resituer correctement que quelques pièces. DE GROOT explique cette différence par la capacité des maîtres à organiser les pièces en groupements plus nombreux et plus importants que les débutants. Les maîtres font en moyenne 7,3 groupements de 3,8 pièces, contre 5,3 groupements de 2,4 pièces pour les débutants.

d) Les expériences sur la réalisation de diagnostics

Les expériences de BORDAGE [BORDAGE 80], ont consisté à présenter à un ensemble de médecins, vingt courts compte-rendus cliniques en leur demandant de rendre un diagnostic et de prendre une décision en un temps limité. Juste après, ce même ensemble de médecins devait répondre par oui ou non à 70 questions portant sur la discipline correspondant aux 20 cas. BORDAGE obtient 40 % d'erreurs de diagnostics pour seulement 5 % d'erreurs aux questions de connaissance pure. Ceci met en évidence les difficultés à utiliser un certain nombre de connaissances, pourtant toutes pertinentes, et que ces connaissances sont intégrées à des mécanismes de traitement d'informations.

e) Les références en psychologie cognitive

Les principaux travaux en psychologie cognitive indiquent que le nombre maximal d'items qu'une personne peut manipuler en même temps est de l'ordre de 7 items.

Dans Pople [POPLE 82], on retrouve cette idée de limitation du nombre de décisions alternatives possibles, et que même si l'informatique permet une augmentation du nombre d'items manipulés ou d'hypothèses envisagées, cette augmentation n'augmentera pas obligatoirement leur pertinence.

2) Un raisonnement sous-tendu par un modèle cognitif

Ces différents travaux permettent d'avoir une idée de la démarche suivie par un expert.

L'expert va plus vite dans la résolution d'un problème ou la recherche d'une solution parce qu'il sait sélectionner d'emblée les informations pertinentes, et contracter le raisonnement en se rapportant à des situations qu'il a déjà traitées. L'activité d'un expert se décompose schématiquement en deux phases, la première étant l'interprétation des signes trouvés et l'évaluation des hypothèses.

L'observation d'un expert indique que ses premières démarches sont d'abord liées à la prise en compte rapide des faits jugés les plus importants par son expérience. La focalisation sur un nombre minimal d'informations est la caractéristique de sa première phase d'analyse d'un problème. Dans cette phase, qui correspond à l'observation de la situation et à son interprétation, l'expert fait souvent référence aux notions de contextes, de stéréotypes, de classifications.

Cette notion de contexte sera détaillée plus loin.

Une fois que les caractéristiques globales du problème sont identifiées par un expert, il passe à un second type d'activité, qui correspond à la mise en relation de la situation avec les objectifs pour rechercher une solution adaptée au problème. Dans cette activité, l'expert envisage des actions élémentaires, puis un nombre limité de plans d'action ou de schémas de décision.

D) Deux notions centrales : le contexte et le schéma de décision.

1) La notion de contexte

Cette notion se retrouve dans pratiquement tous les types d'expertise, et elle apparaît particulièrement importante dès qu'il y a une démarche inductive.

Un contexte peut être ainsi défini par référence aux faits, aux entités, aux conditions à prendre en compte pour décrire la situation dans laquelle une liste d'hypothèses va être construite.

Les systèmes experts de diagnostic en médecine, comme MYCIN [SHORTLIFFE 76] formulent des hypothèses diagnostiques à partir de l'état général du patient et de quelques informations traitées prioritairement .

Les expériences faites avec les joueurs d'échec, montrent que ceux-ci, devant une situation donnée, l'analyse globalement en la rapportant à un cas type qu'ils ont déjà étudié.

D'une manière générale, un contexte est construit par la sélection d'un nombre restreint d'informations traitées prioritairement, et les schémas d'identification et d'interprétation stéréotypés qu'il utilise, indiquent la façon dont il prend en compte l'information disponible.

Cette notion de sélection d'informations est une notion clef, car, lorsque l'expert fait face à une situation, il dispose en général d'un très grand nombre d'informations.

Vogel [VOGEL 88] remarque d'ailleurs que la capacité de déchiffrement d'un expert est fonction de la richesse et de la variété des contextes dont il dispose, et qui activent un nombre limité de schémas d'interprétation synthétiques et puissants.

2) La notion de schéma de décision

Une fois qu'un contexte a été identifié, l'expert se remémore l'ensemble des conduites possibles dans ce contexte, conduites que son expérience du domaine a déjà fait analyser dans des situations proches. La trame de son raisonnement est déjà connue, et son travail consiste principalement à adapter son raisonnement aux caractéristiques propres de la situation présente.

L'expert a souvent mis plusieurs années pour construire ces schémas à partir de l'expérience du domaine qu'il a acquise, et il sait parfaitement les utiliser.

Les schémas de décision peuvent être décrits comme des procédures standardisées d'enchaînement d'actions élémentaires.

F) Les conséquences sur les caractéristiques d'un modèle d'expert

Une des conséquences est que les différents modèles d'expert sont généralement syntaxiquement pauvres mais sémantiquement puissants, et caractérisables par un nombre limité d'éléments.

D'après Vogel [VOGEL 88], l'expert manipule des objets concrets, utilise un système de description qui lui est propre et communique par stéréotypes. L'expert utilise des formes d'autant plus stéréotypées qu'il

oeuvre dans sa zone de plus grande expertise, et des formes d'autant plus ouvertes qu'il se trouve dans une zone d'expertise en cours de constitution.

Une seconde conséquence est que les inférences produites par un expert sont généralement courtes en chaînage et en temps de réponse. Cela veut dire que l'expertise, comme le sens commun, a pour effet de rapprocher les causes des effets d'un phénomène en contractant le raisonnement, et en levant les ambiguïtés entre les interprétations concurrentes.

Le plus souvent, l'expertise acquise par des années d'expérience, a consisté à regrouper les inférences possibles dans des formes stables, liées d'une part aux caractéristiques observables des phénomènes, et d'autre part à un ensemble de contextes activant le schéma d'interprétation.

II) L'ORGANISATION D'UNE DEMARCHE DE RECUEIL D'EXPERTISE

A) Finalité de la démarche de recueil d'expertise

La construction d'un système-expert commence par une phase d'acquisition de la connaissance. Une bonne maîtrise de cette phase délicate suppose que l'on ait acquis une bonne compréhension des caractéristiques générales des modèles propres aux experts.

D'une manière générale, l'expert travaille en suivant un modèle de raisonnement qu'il s'est construit le plus souvent empiriquement. Tous les modèles d'expert ne sont évidemment pas uniformes, mais une analyse indique que l'on retrouve un certain nombre de points communs.

B) Identification et localisation de l'expertise du domaine

La connaissance sur un domaine particulier est multiforme, en partie explicite et en partie implicite. Elle peut avoir déjà fait l'objet de rapports écrits, de schémas, de brochures de documentation. Un premier travail consiste donc à rassembler toute la connaissance écrite sur un sujet donné, en identifiant à chaque fois les auteurs et en recherchant ce qu'ils sont devenus. En effet, toutes ces personnes ont déjà effectué un travail de recueil d'expertise, et pouvoir obtenir leur coopération est primordial.

C) Le recueil des "grains de connaissances"

1) Une analyse fine du domaine

Echanger ou transmettre une expertise suppose avoir un langage commun. Or la connaissance de tout langage commence par l'apprentissage des mots élémentaires. Il en est de même dans le recueil d'expertise.

La construction d'un modèle d'expert commence donc par l'identification et la représentation des objets utilisés dans le domaine. Le savoir étant d'abord classificatoire, la première phase consiste à décrire les objets, en levant les ambiguïtés sur leur définition, puis à les regrouper dans des classes significatives par rapport au champ de l'expertise.

Cette phase de description statique se prolonge par l'identification des actinomies, c'est à dire l'ensemble des séquences d'actions élémentaires provoquant un changement (d'état, de lieu,...) sur un ou plusieurs objets, avec la description des contraintes et des contextes d'application de ces actions.

2) L'identification des objets, puis des classifications pertinentes

Une taxinomie constitue une énumération, une organisation et une description de catégories conceptuellement pertinentes pour l'expert du domaine considéré. Une bonne classification des éléments d'un domaine de référence consiste à les regrouper en fonction de la nature des relations qu'ils entretiennent entre-eux, en se basant sur les similarités et discontinuités opérationnelles que l'on doit identifier au cours de la prise d'expertise.

Dans la plupart des cas, il est en effet possible de générer plusieurs classifications à partir d'un même ensemble d'objets. Le choix de l'une d'entre-elles doit être basé sur la hiérarchie des critères définissant les discontinuités dans un domaine de référence donné. Par exemple, une pompe présente plusieurs attributs, comme son type (vitesse variable ou fixe), sa marque, sa puissance, son prix, etc...

Dans un problème de recherche de configurations de pompage, il est clair que le prix d'achat de la pompe importe peu. Mais cet attribut serait très important dans un problème d'optimisation d'investissements.

Vogel [VOGEL 88] avance l'idée que la plupart des classifications naturelles possèdent des caractéristiques stables, en termes d'abondance de classes ou sous-classes dans les différents niveaux de l'arbre généré. Donc, si une classification consécutive à une expertise s'écarte des modèles standard, il est recommandé de procéder à une analyse plus fine pour identifier les raisons des écarts, qui peuvent être tout à fait justifiés.

3) L'identification des actions élémentaires

Ces actinomies décrivent l'organisation de séquences d'actions et les combinaisons de changements d'état sur les objets, maintenant associés dans un réseau dynamique qui recoupe les réseaux taxinomiques.

Une action n'existe que si elle est précédée d'un certain nombre de conditions qui la rendent possible ou utile, décrites dans le contexte. D'autre part, toute action élémentaire, et en conséquence, toute séquence d'actions, provoque des changements dans les états des objets dont il convient de gérer l'enchaînement.

D) La construction du modèle cognitif de l'expert :

1) Le rôle d'un modèle cognitif :

Si l'on se réfère à VOGEL [VOGEL 88], un modèle cognitif a pour but de "prendre en compte réellement les structures de représentation qui permettent à l'expert de tamiser ses perceptions, de les orienter, de substituer des catégories discrètes à des grandeurs continues, de regrouper des symptômes dans des stéréotypes, d'organiser ces stéréotypes en structures de connaissance, et finalement de mobiliser cette culture pour interpréter et prévoir".

La connaissance d'un domaine ne peut être réduite à une série de règles, plus ou moins indépendantes. Dans la plupart des systèmes à base de connaissance traitant de cas non triviaux, l'ensemble des règles ne constitue généralement qu'une partie du modèle reproduisant la démarche de l'expert.

Une des méthodes de construction d'un modèle consiste à relier les classifications et les séquences d'actions ou de décisions à des contextes qui deviennent des schémas d'interprétation des informations initiales conduisant à des plans types.

2) L'identification de schémas d'interprétation ou contextes :

La collecte des schémas d'interprétation s'opère de différentes manières :

- relevé des modèles physiques connus de l'expert comportant la définition des objets utilisés, la description des manifestations observables et l'identification des phénomènes physiques qui en sont la cause,

- identification des relations entre les différents événements prévisibles et entre les différents modèles de comportement des objets étudiés,

- rattachement de chacun des contextes identifiés à un niveau de généralité de la taxinomie des objets de référence,

- identification des procédés permettant à l'expert de lever les ambiguïtés au moment du choix d'une interprétation pour un phénomène donné,

- identification des procédés permettant à l'expert de contrôler la cohérence de son interprétation.

3) La définition des plans types

Un bon expert recommence rarement un raisonnement dans son ensemble. Il se sert des enchaînements logiques d'actions qu'il a contractés dans des plans types. Un plan type correspond à un enchaînement défini d'actions élémentaires sans qu'il y ait de ruptures ou de points de décision entre les différentes actions.

Pour définir ce type de plan, le cogniticien doit identifier les séquences logiques d'actions élémentaires que l'expert utilise le plus fréquemment.

4) La mise en relation des contextes et des plans d'intention

Une fois que le contexte du cas traité a été identifié, le raisonnement pour atteindre un but est alors structuré par un plan d'intention.

Sa démarche consiste alors à intégrer dans son plan d'intention les plans types qu'il connaît et qui se réfèrent le mieux à la situation traitée. Cela signifie que, généralement, l'expert fait correspondre à un contexte donné, un ensemble limité de plans types qui peuvent s'y référer, et sur lesquels il peut prendre appui pour accélérer sa démarche. Son attention est alors libérée des enchaînements connus, et il peut se concentrer sur les enchaînements nouveaux et spécifiques du cas traité.

D) Présentation de quelques techniques de recueil d'expertise

Plusieurs auteurs [HART 88], [BONNET 87], ont décrit un certain nombre de techniques pour recueillir l'expertise. On retrouve régulièrement celles qui sont décrites dans les paragraphes suivants .

1) L'observation ouverte du comportement de l'expert

Cette technique consiste à examiner l'expert pendant la réalisation de la tâche que le système-expert doit reproduire.

Une telle observation permet, en général, d'identifier le scénario suivi et la décomposition de la tâche complète en tâches élémentaires, mais à moins d'interroger l'expert pendant le déroulement de son expertise, il est pratiquement impossible de découvrir les raisons du cheminement suivi.

2) L'entretien avec l'expert (ou la verbalisation provoquée)

La prise d'expertise par entretien ne se limite pas à une écoute attentive du discours de l'expert reconnu du domaine. Tout entretien

suppose une préparation dont le minimum est l'assimilation du vocabulaire de base du domaine.

La grande difficulté vient du fait que le savoir de l'expert est pragmatique et empirique, et que l'expert ne rend sa démarche explicite que dans de très rares cas. Ce que demande le cogniticien, c'est-à-dire détailler finement le cheminement d'un raisonnement, est inhabituel et peut paraître pénible à l'expert.

Quand l'expertise fait appel à plusieurs domaines, donc à plusieurs experts différents, il est fortement conseillé de les interviewer séparément d'abord, avant de les réunir ensemble.

Tout entretien avec un expert doit être analysé, et les informations extraites lui être présentées pour vérification.

3) Les études de cas historiques

L'analyse des cas précédemment traités par les experts est une source importante d'informations. D'autre part, leur collecte permet de constituer une base de cas, qui sera tout à fait utile lors de l'ajustement du raisonnement.

L'étude de cas permet une mise à plat hors du contexte d'urgence de la plupart des expertises, et donc une recherche beaucoup plus exhaustive des éléments qui rendent explicite la démarche de l'expert.

4) Les questionnaires

Dans le cas où il y a plusieurs experts, dont on souhaite prendre l'avis, la réalisation d'un questionnaire peut permettre de dégager les points de consensus et de désaccord, et donc de reprendre des entretiens cherchant à préciser les causes des désaccords, et donc d'approfondir l'expertise.

Il se peut aussi que les désaccords révèlent des contradictions entre les experts, et qu'il soit nécessaire d'exclure des avis ou alors de limiter l'expertise du système à la zone de consensus.

5) L'expérimentation d'essais de formalisation partielle de la connaissance

Dans le cas où l'expert aura, après la construction du système-expert, à maintenir les bases de connaissances ; il est indispensable de familiariser l'expert avec les différentes possibilités de représentation de la connaissance et de l'intégrer au processus de choix.

C'est pendant cette phase qu'il faut envisager l'utilisation ou non des objets, suivant la structuration des données et la nécessité de transférer des informations par les mécanismes d'héritage. Un raisonnement avec des données incomplètes demande de prévoir des valeurs par défaut, et l'utilisation de frames peut être tout à fait appropriée. Si le raisonnement fait appel à des relations entre objets, la logique formelle constitue un formalisme puissant pour manipuler ces relations. De même que l'utilisation d'objets actifs est tout à fait adaptée à certains problèmes d'ordonnancement, où un objet doit pouvoir faire appel à des procédures ou à des procédures définies dans d'autres objets.

III) L'ACQUISITION DE L'EXPERTISE DANS L'APPLICATION "PILOTE"

A) Les grandes lignes de la démarche suivie

La démarche suivie dans la prise d'expertise consistait à vérifier que l'on pouvait retrouver les types d'activité d'observation-interprétation et de schémas de décision.

Il s'agissait donc d'orienter la démarche de prise d'expertise d'une part, pour rechercher les informations prioritaires qui pouvaient servir à définir et délimiter des contextes d'exploitation, d'autre part pour identifier les actions élémentaires et leur enchaînement intervenant dans la décision.

Comme il a été indiqué au chapitre précédent, toute prise d'expertise commence par sa localisation, puis fait appel aux différentes techniques décrites dans le chapitre précédent.

On est aussi dans un domaine où plusieurs experts travaillent et prennent à tour de rôle des décisions dans le champ d'expertise. Une ligne directrice a donc été d'homogénéiser l'expertise de façon à ce que l'ensemble des intervenants s'y retrouve. Cela a conduit à élaborer des questionnaires, pour tendre vers un consensus.

B) La localisation de l'expertise :

Le premier expert du projet PILOTE a été le directeur technique de l'usine, qui a comme fonction de définir les objectifs de remplissage des différents réservoirs et les conditions d'utilisation des pompes vis à vis du contrat EDF.

Les chefs de poste ont constitué un groupe de 8 experts (5 titulaires et 3 suppléants), car ce sont eux qui ont la responsabilité de définir des stratégies de pompage à partir des consignes générales fixées par le directeur technique.

Ces 8 personnes ont l'expérience concrète de mise en oeuvre de consignes par des choix appropriés de configuration de pompage et de répartition d'eau sur des intervalles horaires donnés.

La plupart d'entre-eux n'avaient pas formalisé cette connaissance. Quelques-uns utilisaient des petits programmes de calculs de dégrossissage, dont l'idée a été reprise dans la méthodologie.

Par contre, il n'y avait aucun document d'aide à la conduite d'usine.

Une fois que l'expertise a été localisée, le recueil proprement dit s'est déroulé en 5 phases successives :

- entretiens avec le directeur technique,
- observation directe sur le poste de commande,
- synthèse des observations et élaboration de questionnaires avec le directeur, le directeur technique, et des ingénieurs des services centraux suivants le dossier,
- remplissage des questionnaires par les chefs de poste,
- analyse des réponses et synthèse avec les acteurs précédents.

C) Les entretiens préliminaires avec le directeur technique :

1) Les objectifs de cette phase

Le premier objectif de cette série d'entretiens était de définir les termes du domaine en essayant de limiter les ambiguïtés, les objectifs attendus d'un schéma de raisonnement.

La description de cet univers a pour finalité de cerner tout ce à quoi le système pourra faire appel, et donc de délimiter le champ dans lequel il opérera.

Ensuite, il s'agissait de vérifier que dans le cas de la conduite d'usine, l'expert utilisait aussi un raisonnement basé sur l'identification de contextes (les principaux cas de figure), et leur utilisation pour s'orienter vers des schémas décisionnels types. La prise d'expertise a donc compris une phase de définition de la structure de tels schémas.

Le dernier objectif était de définir des critères d'évaluation pour comparer entre-elles des stratégies de pompage..

2) La définition des termes du domaine d'expertise

Les termes du domaine comprennent les objets physiques et les grandeurs manipulées.

Les objets physiques du domaine sont :

- les groupements de pompes (ancienne et nouvelle élévatoire),
- les pompes elles-mêmes (petites, grosses, super),
- les réservoirs (élévation, effacement)

Les grandeurs manipulées sont :

- pression de refoulement,
- hauteur manométrique,
- niveaux de réservoir.

Des termes ont été introduits dans le projet, comme, par exemple, stratégies de pompage.

3) Les objectifs visés par l'expert

L'objectif du directeur technique était de définir des configurations de pompage à un horizon de 24 heures.

4) L'identification des principaux cas de figure

La partie de l'entretien consistant à rechercher les causes influant sur les consignes a conduit à définir la notion de contexte d'exploitation. Celui-ci est déterminé dans ce domaine d'expertise, toujours à partir d'un petit nombre d'informations, connues initialement et traitées prioritairement dans les règles d'élaboration des consignes.

Ces informations sont :

- le type de journée vis à vis du contrat EDF,
- l'ordre de grandeur du volume d'eau à produire,
- l'indisponibilité de réservoirs dans ou hors de l'usine,
- l'état de fonctionnement de l'usine "jumelle" de Choisy.

5) La structure d'un schéma décisionnel

Le second objectif était de décrire les types d'indications données dans une consigne de conduite d'usine. Une consigne globale est en fait composée de niveaux à atteindre à des heures données.

Donc une consigne implique une structuration de la journée, qui correspond à un découpage implicite de la journée en périodes de planification. La journée est décomposée en intervalles au bout desquels on fait des bilans partiels provisoires, et ceci est identifiable comme la décomposition du problème initial en sous-problèmes.

6) Les critères d'évaluation des stratégies de conduite d'usine

Le premier jeu de critères défini par le directeur technique de l'usine pour comparer des stratégies de pompage concurrentes a été le suivant :

- risque d'insuffisance d'approvisionnement en eau

- risque de dépassement des limites inférieure ou supérieure des réservoirs internes,
- risque de dépassement du contrat EdF,
- risque de casse sur le réseau.

Un critère d'un autre type, la régularité du régime nourricier, a été présenté pour prendre en compte la qualité de l'eau, sur le plan hydraulique.

Le nombre de changements de configuration élévatoire a été présenté pour prendre en compte des problèmes de maintenance dans l'évaluation des stratégies, en même temps que le risque de casse sur le réseau.

Le dernier critère est économique puisqu'il s'agit du coût énergétique du pompage élévatoire. La base de la comparaison sera le coût moyen pour élever 1 m³ d'eau pendant la période de planification.

D) L'observation sur le poste de commande :

1) Les objectifs de la phase d'observation

Suite à la série d'entretiens avec le directeur technique, il importait, d'une part de voir comment les consignes étaient mises en application sur le poste de commande, et d'autre part de rendre explicite l'ensemble de la démarche de choix d'une configuration de pompage.

2) Une identification des informations structurantes des contextes

L'observation a aussi mis en évidence que les chefs de poste n'envisageaient qu'un petit nombre de stratégies potentielles de pompage en fonction des contextes d'exploitation. De plus ces contextes sont bien identifiés à partir des mêmes informations que celles qui avaient été données par le directeur technique.

3) Mise à plat de schémas décisionnels

L'observation attentive a fait apparaître que les chefs de poste utilisaient bien des objectifs intermédiaires, fonctionnant comme des points de repère, même s'ils ne devaient définir une stratégie de pompage que pour 8 heures. En fait les consignes de niveaux successifs à viser, intègrent la notion de continuité du pompage sur 24 heures. Il a donc été possible d'identifier un certain nombre d'heuristiques de structuration de la journée, présentées plus loin.

Il est aussi apparu que les chefs de poste avaient construit des modèles empiriques de consommation d'eau, et qu'ils se servaient des consommations prévisionnelles dans la phase de décision.

La démarche des chefs des postes peut se caractériser par 2 phases. Dans la première, ils envisagent un nombre limité de stratégies de pompages, qu'ils étudient ensuite plus finement à l'aide de quelques calculs sommaires incluant des prévisions de consommation. Un deuxième ensemble d'heuristiques, utilisées pour restreindre le nombre de stratégies envisagées, a pu être explicité.

4) Les heuristiques de structuration de la journée

a/ La finalité de ces heuristiques :

Une fois que le type de jour et la situation d'exploitation de l'usine sont connus, la journée peut être décomposée en périodes pendant lesquelles les valeurs potentiellement intéressantes des variables de commande sont limitées.

Ces heuristiques structurent la journée par des points de repère, en leur associant des objectifs intermédiaires, dont on sait, par expérience, qu'ils sont les points de passage recommandés pour aboutir aux objectifs terminaux.

Elles déterminent aussi, l'écart à l'objectif à partir duquel il est très vraisemblable, mais pas certain, qu'une stratégie conduira plus tard à une impasse.

b/ Les découpages de journée en plages horaires :

Les journées en situation normale, sont plus ou moins fortement structurées par un certain nombre d'éléments, liés aux contraintes, aux nécessités de la production, et à la vie de l'usine. Deux exemples sont donnés.

b.1/ Pour les jours ouvrables d'hiver, la structuration est très nette et définie par les termes du contrat C.G.E.-E.d.F., qui impose l'arrêt des élévatoires de 9 h à 11 h et de 18 h à 20 h. Cette contrainte définit par elle-même les points de repère : 11h et 20h et donc les 3 plages de la journée. (6 h - 11 h), (11 h - 20 h) et (20 h - 6 h).

b.2/ Pour les samedis et dimanches d'hiver, où la tarification et la disponibilité en énergie électrique restent identiques quelle que soit l'heure, la structuration est beaucoup plus faible et prend appui sur le profil de consommation horaire. La journée se décomposera en deux plages : 6h - 22h

correspondant au pompage de jour, et 22h - 6h à celui de nuit. Il en est de même pour les jours de printemps et d'automne. L'été n'a pas été pris en compte.

5) Les heuristiques de restriction des stratégies envisageables

L'observation de la conduite d'usine en temps réel met rapidement en évidence le fait que les chefs de poste n'envisagent, sur une plage horaire déterminée, que quelques stratégies de pompage.

Un très grand nombre de stratégies ne seront pas examinées parce que l'expérience acquise indique, a priori, qu'elles conduiront à une situation inacceptable.

Ce groupe d'heuristiques a donc pour objet, d'éliminer d'emblée, ou après un calcul sommaire, toutes les solutions dénuées d'intérêt dans une situation donnée. Ces heuristiques sont décomposables en trois catégories définissables par leur fonction :

- réduction du nombre de configurations sur un pas de temps,
- réduction du nombre de sous-plages dans la plage horaire,
- élimination de stratégies à partir d'un calcul sommaire.

a/ Heuristiques limitant le nombre de configurations explorées sur un pas de temps

Il a été vu précédemment que, pour chaque pas de temps, 62 configurations de pompes étaient envisageables. Ce nombre peut être considérablement réduit dans une situation d'exploitation normale, à partir des considérations suivantes :

a.1/ une recommandation d'utilisation des deux réservoirs d'effacement

L'utilisation des deux réservoirs d'effacement minimise le risque de changement de régime nourricier. Aussi, en situation normale, les stratégies utilisant uniquement les pompes de l'ancienne ou de la nouvelle élévatoire ne seront pas examinées.

Cette heuristique élimine 22 configurations.

a.2/ Minimisation des pertes de charges à l'aspiration :

Compte-tenu de la disposition des réservoirs d'effacement alimentant l'ancienne élévatoire, les pertes de charge deviennent très importantes quand trois pompes ou plus de type E₁-E₆ sont simultanément en fonctionnement.

Cette heuristique élimine 12 nouvelles configurations.

a.3/ Dépassement de contrat EdF :

Les configurations comprenant deux super pompes et au moins 3 pompes de types E₁-E₆ consommeront une puissance électrique supérieure à la puissance disponible dans le contrat EdF.

Cette heuristique rejette 12 autres configurations.

a.4/ Bilan de l'utilisation de ces heuristiques :

Ces trois heuristiques ramènent à 16 le nombre de configurations envisagées sur un pas de temps horaire.

b/ Heuristiques limitant le nombre de sous-plages :

Dans une plage horaire durant plusieurs heures, il est possible d'examiner des stratégies présentant des changements de configurations à chaque pas de temps.

Mais l'expérience de la conduite d'usine indique qu'il est souhaitable de minimiser le nombre de changements de configurations.

En effet tout arrêt ou mise en route de pompes induit des oscillations de pression dans le réseau, qui présentent, particulièrement à pression élevée un risque de casse. Tout démarrage de groupe appelle un courant de grande intensité qui est un important facteur d'usure du moteur électrique des pompes.

Une évaluation du coût de la maintenance des pompes en relation avec leur nombre de démarrages, a conduit à définir comme heuristique, la limitation du nombre de changements de configurations, à un par plage horaire. Les configurations avant et après l'heure de changement sont alors définies en relation avec l'objectif intermédiaire correspondant à cette heure de changement.

Pour les plages horaires des après-midis des jours ouvrables d'hiver, entre 11 h et 20 h, ou les jours de week-end, l'heure de ce changement a été définie en relation avec le profil de consommation en eau. Un changement sera envisagé à 14 heures, ce qui correspond à la fin de la pointe de consommation du matin et au début de la période étale de l'après-midi.

Pour les nuits des jours ouvrables d'hiver, les heures de changement envisageables sont 22, 23, 24 ou 1 h, selon que l'on veuille commencer plus ou moins tôt le remplissage des réservoirs en jouant sur les tarifications EdF de nuit.

6) La prise en compte des critères de choix des stratégies

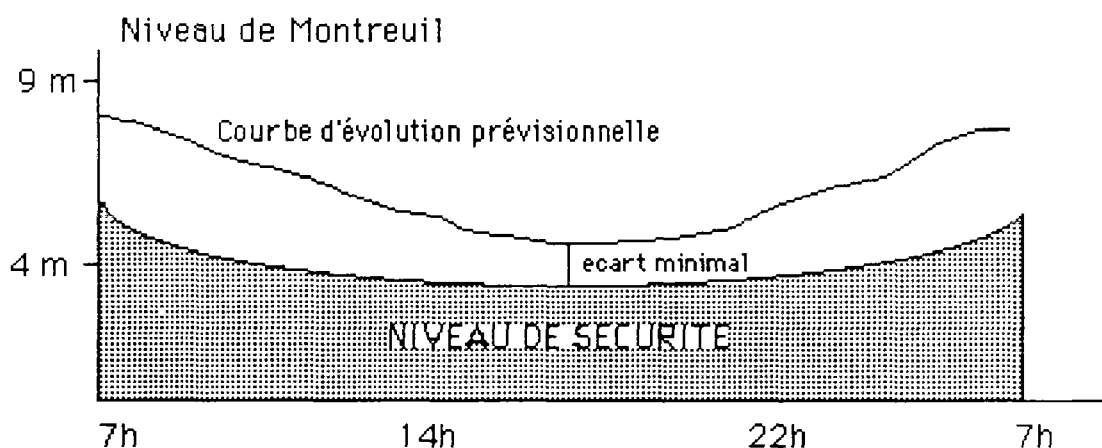
Dans les paragraphes précédents, il a été présenté des heuristiques limitant le nombre de configurations envisagées. Ce nombre n'étant pas réduit à 1, il était nécessaire d'avoir des éléments de comparaison de différentes stratégies.

Les entretiens avec les chefs de poste pendant la phase d'observation ont permis de décrire plus précisément un ensemble de 7 critères, et les relier à des valeurs numériques d'évolutions prévisionnelles.

a) La sécurité en approvisionnement

La sécurité en approvisionnement peut être évaluée à l'aide d'une courbe limite de remplissage des réserves, en calculant heure par heure, la distance par rapport à cette courbe, et non pas seulement la distance à l'objectif final.

Le descripteur retenu pour représenter cette notion de sécurité en approvisionnement est la distance minimale au dessus de la courbe limite.



b) La sécurité vue du point de vue électrique :

Le descripteur retenu pour représenter cette notion de sécurité vis à vis du contrat EDF est l'écart minimal entre les puissances appelée et autorisée.

Dans le logiciel PILOTE, la puissance appelée correspond à la somme

- de la puissance appelée par l'élévatoire et la nourricière,
- de la puissance appelée par les ozoneurs,
- de la puissance appelée par le reste de l'usine (décanteurs, DEA...) donnée sous forme de valeurs moyennes (talon).

La sécurité électrique correspond donc à la marge à prendre par rapport au contrat EDF, en travaillant avec un nombre entier d'unités d'intégration.

c) Le risque de rupture sur le réseau

L'évaluation des stratégies sera basée sur la valeur maximale prise par la pression de refoulement pendant la période de planification considérée.

L'introduction de ce troisième descripteur sert à prendre en compte les risques de rupture ou d'usure des canalisations sur le réseau.

d) Le coût énergétique de la solution :

Le tarif auquel EDF vend l'électricité varie suivant l'heure de la journée et de la saison.

Par exemple en automne, les tarifs sont les suivants :

- jour : $c = 0,566 \text{ F/kWh}$.
- nuit : $c = 0,448 \text{ F/kWh}$.

Comme la simulation d'une stratégie donne une évaluation de la puissance appelée par l'usine pour chacune des 24 heures, le coût de cette stratégie est alors donné par l'équation :

$$C = \sum c_i \cdot p_i$$

Ce coût est ensuite divisé par le volume pompé pour avoir une base de comparaison commune avec l'ensemble des autres stratégies. Le coût de la stratégie prend implicitement en compte les rendements des pompes, puisque la puissance consommée est calculée à partir des courbes des pompes.

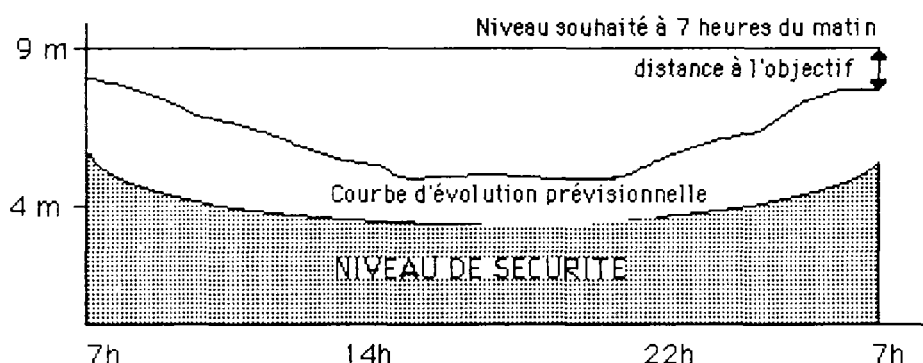
Le coût moyen du pompage d'un mètre-cube pour une journée est de l'ordre de 10,5 centimes.

e) La distance à l'objectif fixé :

Le système PILOTE fixe des objectifs de remplissage du réservoir de Montreuil pour le lendemain matin, 6 heures. L'idée étant de commencer la journée du lendemain avec des réserves suffisantes.

Aussi est-il possible de mesurer l'écart entre le niveau prévisionnel que l'on aurait en suivant la stratégie et le niveau fixé.

Niveau de Montreuil



f) Régularité du régime nourricier

La régularité d'un régime comprend deux aspects : le nombre et l'importance des changements. Mais c'est surtout l'importance du changement de régime qui est pénalisante.

8) Nombre de changements de régime élévatoire

Pour l'usine élévatoire, ce critère prendra en compte le nombre de changements de régime dans les périodes de planification, qui ne peut être que 0, 1 ou 2. Les stratégies présentant plus de 2 changements ne sont pas étudiées.

E) L'élaboration de questionnaires

1) Les objectifs de la phase

La phase d'observation a mis en évidence les points communs structurant la démarche des experts, comme la notion de point de repère, ou le jeu de critères d'évaluation des stratégies de pompage. Mais en même temps, cette phase a fait apparaître des divergences.

Les différences observées entre les comportements des chefs de postes ont conduit à définir une méthode de réduction des divergences. Cette méthode a été inspirée par les expériences menées chez Boeing à Seattle

[BOOSE 86]. BOOSE a défini et répertorié plusieurs degrés d'implication d'un expert vis à vis d'une proposition faisant partie d'une expertise :

- l'expert est entièrement d'accord avec la proposition et la pense évidente.

- l'expert agrée après quelque réflexion. Il n'avait jamais considéré le domaine du problème de cette manière particulière avant,

- l'expert est en désaccord avec la proposition, mais accepte de la considérer comme une exception,

- l'expert rejette catégoriquement la proposition.

Pour bien cerner les divergences, il a été décidé d'élaborer un questionnaire pour rechercher les options sur le découpage de la journée en périodes de planification et sur les niveaux intermédiaires visés en fin de période. Le questionnaire avait pour objet de faire apparaître la valeur la plus consensuelle, pour ensuite aller voir ceux qui s'en écartaient, en essayant de déterminer avec eux, les raisons de ces écarts.

Comme les chefs de poste avaient quelques divergences quant à l'appréciation des intervalles de qualification pour différents critères, et sur l'ordre de prise en compte de ces critères, un second questionnaire a été élaboré, toujours dans la même idée de recherche d'un consensus. L'élaboration de ces questionnaires a été réalisée par le groupe du directeur de l'usine et des ingénieurs de la Compagnie.

2) La recherche d'un consensus sur les découpages et points de repère

Dans cette partie, il s'agissait de d'abord de demander comment la journée devait être découpée, en fonction des différents cas de figure de l'exploitation normale. Une fois la journée découpée, il importait alors de connaître les niveaux de référence des réserves, jugées comme souhaitables par les chefs de poste en fin de période.

3) La recherche d'un consensus sur l'appréciation qualitative des critères

Au moment de l'élaboration des questionnaires, il est apparu nécessaire de distinguer deux notions. La première concerne les perceptions des importances absolues et relatives des critères. La seconde fait référence à des classes de qualification, car les chefs de poste ne raisonne pas en fonction de valeurs numériques, mais en fonction d'une appréciation qualitative des critères.

Dans le questionnaire, les classes ont des noms différents, liés à la grandeur que le critère mesure, comme risquée ou sûre pour une sécurité, ou comme petite ou grande pour une distance. Mais chaque critère est qualifié selon trois catégories, équivalentes à "Bon, moyen, mauvais". L'idée directrice était de faire ressortir les conditions dans lesquelles une stratégie doit être un peu, moyennement ou beaucoup déqualifiée.

F) Le remplissage des questionnaires et l'analyse des réponses

1) Le déroulement de la phase

Les questionnaires ont été distribués aux différents chefs de poste, en insistant sur le fait que l'on raisonnait en avenir.

L'analyse des réponses, complétée par des entretiens individuels, a permis de dégager un consensus sur les différents points abordés :

- le découpage de la journée en périodes de planification,
- les objectifs intermédiaires de fin de période.
- l'ordre hiérarchique des critères,
- les intervalles de qualification de chacun de ces critères.

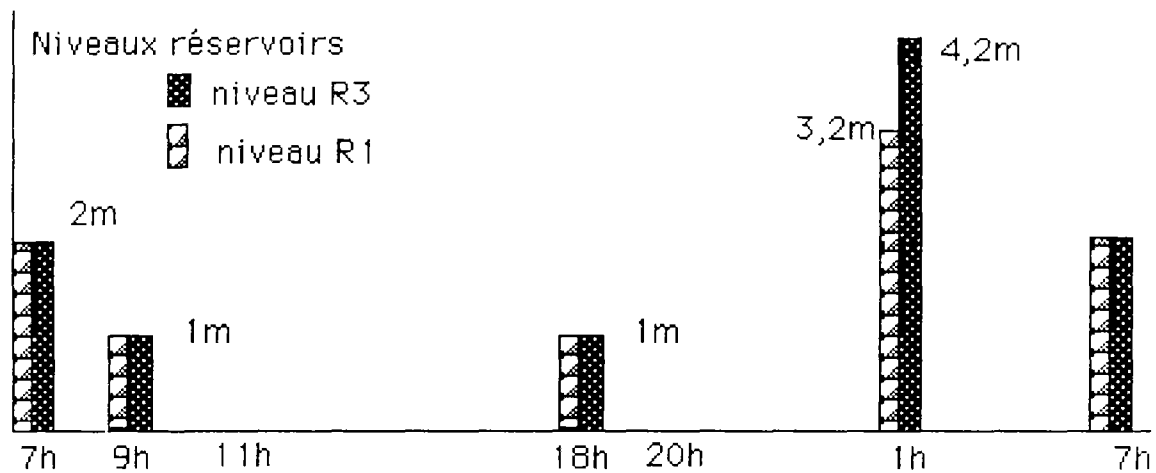
2/ Les découpages de journées et les objectifs servant de points de repère

a) Le découpage des journées

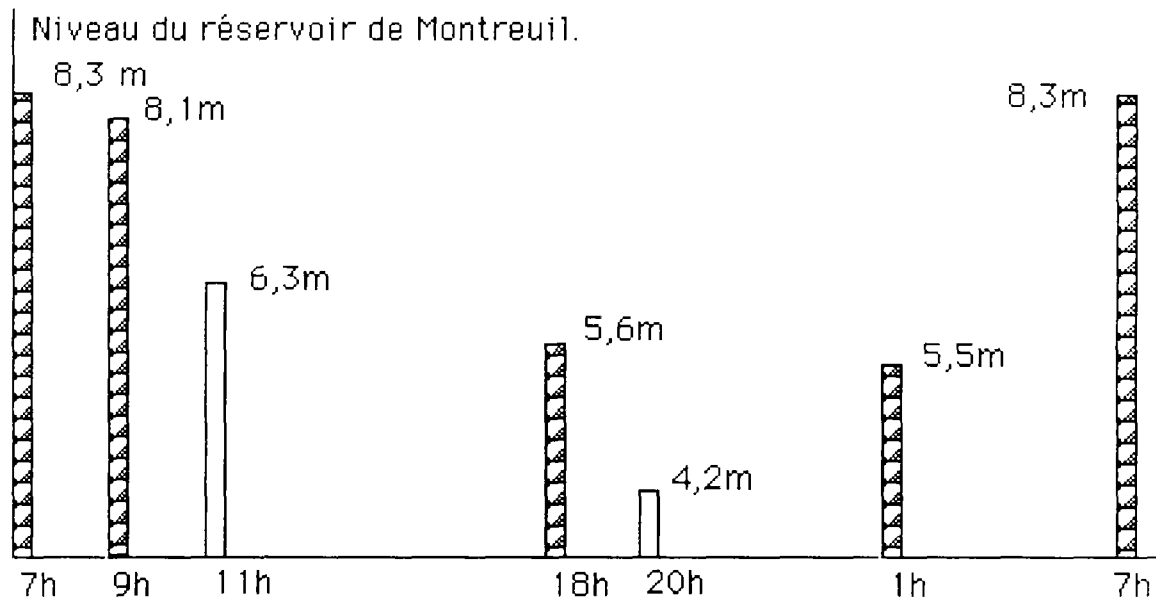
Par exemple, pour les jours ouvrables d'hiver, les journées sont structurées par les périodes de tarification EDF, qui impose pratiquement l'arrêt des élevatoires entre 9 et 11 heures, et 18 et 20 heures. De plus le changement de tarification a lieu à 1 heure du matin.

Les deux points de repère ont été fixés à 9 et 18 heures, soit en début de période d'arrêt de pompage, mais les niveaux de référence à ces heures intègrent la baisse des niveaux des réservoirs qui ont lieu pendant les deux heures suivantes.

b/ Pour les réservoirs d'effacement les jours ouvrables d'hiver



c/ Pour le réservoir de Montreuil en jour ouvrable d'hiver



3) Importance et classes de qualification des critères

a) L'importance

Le questionnaire a permis de définir pour ces critères, la hiérarchie suivante et des classes de qualification :

- niveau minimal de la courbe de remplissage de Montreuil
- marge de sécurité vis-à-vis du contrat EdF
- pression maximale de refoulement
- la distance aux objectifs de remplissage en fin de plage (Montreuil et effacement)
- le nombre de changements de configuration élévatoire

- le coût pour élever 1m^3 d'eau.

b) Les classes de qualification des critères :

Le questionnaire avait aussi pour objet de déterminer les valeurs numériques à partir desquelles on classera les différents critères (sécurité, sensibilité, hmt, distance à l'objectif, etc...) dans l'une des trois catégories équivalentes à : BON, MOYEN, MAUVAIS.

b.1) Pour la sécurité en approvisionnement :

Pour la gestion du réservoir de Montreuil, le questionnaire a fait ressortir les limites de catégories suivantes, pour l'élimination et la déqualification des stratégies.

	3, 5m	3, 65m	4m
-----!-----!-----!-----			
RISQUE MAJEUR	TRES RISQUEE	RISQUEE	SURE

b.2) Pour la sécurité vis à vis du contrat EDF :

Pour définir les marges de sécurité vis à vis du contrat EDF, le questionnaire a fait ressortir les limites de catégories suivantes, Ces valeurs sont les moyennes de l'ensemble des questionnaires, ramenées au multiple immédiatement supérieur du pas d'intégration EDF qui est 90 kW.

	270 kW	450 kW	810 kW
-----!-----!-----!-----			
ZONE DE RISQUE MAJEUR	ZONE DE RISQUE	ZONE DE VIGILANCE	ZONE SURE

b.3) Pression de refoulement maximale :

Pour définir les risques liés aux fortes valeurs des pressions de refoulement, le questionnaire a fait ressortir les limites de catégories suivantes, Ces valeurs sont les moyennes de l'ensemble des questionnaires, ramenées à l'unité immédiatement inférieure et concerne la pression de refoulement de la nouvelle élévatoire.

	100 m	103 m	105 m
-----!-----!-----!-----			
AUCUN PROBLEME	ZONE DE PRESSION ACCEPTABLE	ZONE DE PRESSION FORTE	ZONE DE PRESSION TRES FORTE

b.4) La distance à l'objectif :

Pour analyser l'écart entre le niveau prévisionnel à Montreuil atteint par la stratégie et l'objectif fixé, le questionnaire a fait ressortir les limites de catégories suivantes. Ces valeurs sont les moyennes de l'ensemble des questionnaires, arrondies au dixième de mètre le plus proche.

0,2 m		0,4 m		0,7 m	
-----!-----		-----!-----		-----!-----	
ECART	ECART	ECART	ECART	ECART	ECART
NEGLIGABLE	FAIBLE	APPRECIABLE	APPRECIABLE	IMPORTANT	IMPORTANT

Pour les réservoirs d'effacement, les écarts sont les mêmes.

b.5) La régularité du régime nourricier :

Pour analyser l'importance des variations de régime nourricier, le questionnaire a fait ressortir les limites de catégories suivantes. Ces valeurs sont les moyennes de l'ensemble des questionnaires, exprimées en m³/jour, et arrondies au millier de m³/jour le plus proche.

35000		80000		140000	
-----!-----		-----!-----		-----!-----	
VARIATION	VARIATION	VARIATION	VARIATION	VARIATION	VARIATION
NEGLIGABLE	FAIBLE	IMPORTANTE	IMPORTANTE	TRES IMPORT.	TRES IMPORT.

b.6) Le nombre de changements de configuration élévatoire

Le nombre de changements étant limité à 2 sur une plage horaire, les 3 classes "Bon, moyen, mauvais", correspondent respectivement aux nombres de de changement 0, 1, 2.

b.7) Le coût

Un nombre élevé de calculs de coût sur les périodes de planification ont indiqué que le coût moyen pour élever 1 m-cube d'eau variait de 9 à 12 centimes suivant la période. Mais à l'intérieur d'une même période, les écarts entre stratégies sont limités à 0,4 centimes. Aussi, l'affectation d'une classe de qualification pour un coût sera faite en valeur relative. Le coût minimal correspondra à la classe "bon", le coût maximal à la classe "mauvais", et le ou les 2 coûts intermédiaires à la classe "moyen".

IV) BILAN ET CRITIQUE DE LA DEMARCHE DE PRISE D'EXPERTISE

A) Synthèse de la phase de prise d'expertise

La phase de prise d'expertise a permis de déterminer :

- les règles générales de définition d'un contexte d'exploitation et des consignes qui en résultent.
- la démarche de mise en oeuvre de ces consignes par les chefs de poste, c'est à dire le découpage de la journée avec des points de repère, un modèle d'estimation grossière de la consommation, et la focalisation, par un calcul sommaire et une prise en compte de l'historique, sur un petit nombre de stratégies potentielles.
- l'ensemble des critères de choix entre stratégies potentielles avec leur importance et leur classement.

Il s'est avéré par la suite, comme cela est décrit dans la phase d'évaluation, que la modélisation de la procédure de choix était insuffisante. Avoir définis un ordre hiérarchique de prise en compte des critères et des classes de qualification ne suffit pas à reproduire une démarche, qui, comme on le verra, appréhende les critères en les regroupant en différents niveaux. Un retour auprès des experts s'avère nécessaire, suite à la phase d'évaluation du système.

D'autre part, le classement reflète le point de vue de ceux qui mettent en avant les aspects sécurités, et ce point de vue n'est pas identique avec celui de ceux qui fixent les consignes qui accordent plus d'importance au coût du pompage. Sur ce point aussi un arbitrage est à prévoir.

Il ressort aussi de cette phase quelques lignes directrices qui peuvent servir de références pour aborder une phase de recueil d'expertise à l'occasion d'un nouveau projet.

B) Les remarques issues de l'expérience

La première idée est qu'une phase de recueil ne se déroule pas linéairement, en n'utilisant qu'une seule technique, sur un laps de temps court. Sur l'usine, le recueil de l'expertise a duré près d'un an, et a fait appel à plusieurs techniques, comme l'entretien, l'observation de l'expert, les questionnaires, et mêmes les études de cas.

Il en résulte qu'une démarche de recueil ne peut pas être complètement programmée d'entrée, en se basant sur une seule technique. La collecte doit être constamment réajustée en fonction des éléments précédents.

Il est important d'intégrer la notion de temps dans le projet, particulièrement lorsqu'il y a plusieurs experts et futurs utilisateurs, ce qui nécessite la recherche d'un consensus. Lorsque des positions divergentes apparaissent, il est indispensable de passer voir les différents experts un par un. C'est à l'ingénieur cognitif qu'il appartient de faire le travail de rapprochement des points de vue, en essayant de comprendre les fondements des divergences et de les réduire. La mise en présence des différents experts ne peut être que bénéfique qu'après un profond travail de réduction des divergences.

Une troisième remarque consiste à percevoir la prise d'expertise comme l'explicitation d'un très grand nombre de décisions élémentaires, que l'expert ne formule que rarement, tellement elles lui paraissent évidentes. Le travail de décomposition de l'ensemble des faits et gestes de l'expert, en les reliant à chaque fois aux explications, est souvent long, et perçu comme fastidieux. Il est pourtant indispensable quand on désire implémenter un système.

C) Expertise et système de traitement de l'information

Une première réflexion sur la phase de recueil de l'expertise laisse entendre qu'un schéma décisionnel peut être vu comme un système de traitement des informations.

L'expert en effet sélectionne et analyse un petit nombre d'informations à une étape donnée de son raisonnement. Recueillir une expertise s'apparente souvent à identifier quelles informations sont sélectionnées, et comment elles sont utilisées à chaque phase du raisonnement.

Mais les informations dont dispose initialement l'expert pour traiter le problème ne sont pas toujours suffisantes. Aussi, est-il très fréquent, que cet expert, comme dans le cas du jeu du taquin ou de DENDRAL, génère des informations complémentaires.

Recueillir une expertise consiste donc à construire un système de traitement des informations, comprenant évidemment plusieurs cycles de sélection, traitement et génération d'informations dès que le problème cesse d'être trivial.

Un examen des schémas d'utilisation des informations par les experts, fait apparaître les notions de pertinence et d'accessibilité.

D) Pertinence et accessibilité des informations

1) La pertinence d'une information

La pertinence d'une information est liée à sa capacité d'orientation dans l'arbre de représentation du problème, et au nombre de solutions potentielles qu'elle peut rejeter ou défavoriser fortement à un niveau de la recherche.

Dans tout recueil d'expertise, il apparaît vite que certaines informations ont un poids très important dans l'orientation du raisonnement, que ce soit pour un diagnostic ou une option de planification.

Ainsi, pour les joueurs d'échecs, la localisation du roi et des pièces qui le menacent, vont orienter l'analyse des positions des autres pièces, la perception de l'ensemble de la situation, et en conséquence, la recherche des solutions.

De même, dans l'usine, la classe du tirage et la période de tarification EDF, sont des éléments structurants qui délimitent, d'emblée, les principales options de planification.

Un travail de recherche sur la pertinence des informations s'articule bien avec l'identification des éléments structurants des contextes. Cela donne la trame d'un travail de recherche sur la mise en place de protocoles de recherche de ces éléments structurants.

2) L'accessibilité d'une information

Il est aussi évident que tout processus de traitement des informations prend en compte leur facilité d'acquisition.

La difficulté d'obtenir une information spécifique, peut venir de sa dilution dans un grand nombre d'informations. Elle peut aussi être due aux traitements nécessaires d'un ensemble d'informations élémentaires brutes.

3) Une oscillation entre pertinence et accessibilité

L'observation de la façon dont un expert utilise les informations, montre que toutes les informations immédiatement disponibles ne sont pas utilisées. De même, l'expert ne recherche pas systématiquement les informations les plus pertinentes, quand elles sont difficilement accessibles.

Sur l'usine, les experts utilisent les informations par niveau d'accessibilité. A un niveau donné, les informations sont prises en compte selon leur degré de pertinence. Comme on peut observer régulièrement que tout ce qui est disponible à un palier d'accessibilité donné, n'est pas utilisé, on peut formuler l'hypothèse qu'il existe, à chaque niveau, un seuil de pertinence des informations. Identifier les informations qui sont en-deça de ce seuil, est une opération qui a pour objet de mieux focaliser le système sur les informations importantes.

Ce seuil déclenche la recherche d'informations complémentaires. Une autre observation est qu'une telle recherche apporte rarement une seule information, mais tout un ensemble dont le degré de pertinence varie. C'est bien pour cette raison qu'il est important de mettre rapidement de côté, les informations de faible intérêt.

4) La mise en place de protocoles de recherche

Demander à l'expert quelles informations il utilise, n'est pas évident. Par contre, il peut être beaucoup plus illustratif pour la méthodologie de rechercher quelles sont les informations dont l'absence est la plus pénalisante, pour un raisonnement ou la formulation d'un diagnostic.

Cela peut être envisagé par la mise en place d'un protocole, basé sur des études de cas, formés d'un ensemble d'informations incomplètes. Le protocole expérimental consisterait à analyser, au cas par cas, les demandes d'informations complémentaires formulées par l'expert, et la dégradation de la qualité de son raisonnement, en leur absence.

L'identification des cycles de traitement des informations constituerait la trame de base de la construction du modèle de l'expert.

CHAPITRE IV

DE L'INTEGRATION DE L'EXPERTISE A LA CONCEPTION DU SYSTEME

* * * * *

I) LA CONSTRUCTION D'UNE METHODOLOGIE ADAPTEE AU PROBLEME

A) Le passage du modèle de l'expert à la base de connaissances :

Une fois que les principaux éléments du modèle qui représente le mieux la démarche de l'expert sont identifiés et décrits, le cognicien peut construire la base de connaissances en utilisant les éléments explicites décrits dans le modèle de l'expert, et en les intégrant dans les stratégies de résolution selon l'état de l'art en méthodologie et en moyens de calcul.

En effet, la construction d'une base de connaissances ne se limite pas à reproduire totalement la démarche de l'expert, mais à la rendre suffisamment explicite pour pouvoir l'intégrer en la combinant aux moyens de calcul et de raisonnement implémentables.

Pour reprendre un exemple souvent cité, il est évident que, si l'étude du vol des oiseaux a largement contribué à l'amélioration des connaissances sur les lois de l'aérodynamisme, ce n'est pas en reproduisant exactement le mode de propulsion des oiseaux que les ingénieurs aéronautiques ont pu faire voler les avions. De même, l'éplucheuse de légumes a été conçue autour d'une centrifugeuse et non à partir de bras articulés.

La phase active et itérative d'acquisition de l'expertise auprès de l'expert ou des différentes sources de documentation, se prolonge par une phase d'intégration des différents éléments, et de conceptualisation.

La phase suivante d'opérationnalisation de la connaissance, qui aboutira donc à la construction des bases de connaissance, en suivant le formalisme de représentation et la méthode de résolution retenus, est décrite dans le chapitre suivant.

Dans le cas du problème traité, la phase de prise d'expertise a servi à formaliser l'ensemble des étapes conduisant à la détermination d'une stratégie de pompage. Il s'agit maintenant de construire un modèle cognitif en analysant l'ensemble des objets du domaine manipulés, en identifiant les acteurs, et le champ décisionnel dans lequel ils interviennent.

BUCHANAN [BUCHANAN 82], souligne trois critères fondamentaux à prendre en compte pour envisager une représentation de la connaissance dans un système expert :

- l'extensibilité,
- la simplicité,
- une connaissance explicite.

C'est en se référant à ces trois axes que la phase de conception a été abordée.

B) Identification des caractéristiques du problème posé

Le recueil de l'expertise et la reproduction de la démarche de l'expert ont fait apparaître un certain nombre de caractéristiques du problème.

1) Une recherche de solutions dans un espace non prédéfini

Une des caractéristiques de ce problème de planification, faite en avenir incertain par définition, est que la solution est recherchée dans un espace non prédéfini qu'il faut construire et délimiter lors de chaque session. Il en résulte qu'une recherche exhaustive n'est envisageable qu'au prix d'un calcul incompatible avec les moyens mis en oeuvre et qu'il n'existe pas à priori de solution optimale.

2) La nécessité de construire un modèle prévisionnel

Comme dans tout problème de planification, il a été nécessaire de construire un modèle de prévision de consommation, dont les fondements ont été donnés dans le chapitre I. Ce point est important, car il signifie que les évaluations faites par le modèle vont fortement influencer les décisions de gestion.

3) Une possibilité de décomposer le problème en sous-problèmes

Il est apparu lors du recueil d'expertise que le problème était décomposable en sous-problèmes, c'est à dire que la planification sur une période de 24 heures pouvait être vue comme une succession de planifications sur des périodes plus courtes, à condition de fixer des objectifs intermédiaires. Ces objectifs existent bien dans ce cas, puisqu'ils servent de points de repère pour le directeur technique et les chefs de poste.

4) L'existence d'un modèle de simulation

Il a été vu dans l'analyse de l'existant que l'on disposait de modèle de prévision et de simulation qui peuvent être vus comme des instruments de création d'informations, sous forme d'un ensemble de données numériques indiquant les évolutions prévisionnelles.

5) La possibilité de corriger une décision erronée en temps utile

D'autre part, le process de traitement de l'eau étant relativement long, les évolutions sont lentes et repérables. En conséquence, une stratégie erronée peut être modifiée avant qu'elle ne conduise à une situation dangereuse du point de vue de l'exploitation. Mais il est évident que des remises en cause de stratégies provoquent des perturbations dans la régularité de la conduite d'usine, et qu'un des objectifs du système est de les éviter. Il n'en demeure pas moins que toutes les contraintes liées au temps réel, ne s'appliquent pas à ce système.

6) Un ensemble restreint de données initiales

Le nombre d'informations pertinentes et disponibles au moment de la recherche d'une stratégie de pompage est limité à une quinzaine. Cet ensemble est d'autant plus faible, qu'une partie de ces informations peut être déduite implicitement, à partir de la date, par exemple.

7) Un ensemble de contraintes strictes

La gestion hydraulique impose des contraintes strictes qui seront autant d'éléments pour rejeter les stratégies qui ne les respectent pas. Ces contraintes ont été décrites au chapitre précédent, et une bonne stratégie de résolution doit s'en servir, pour restreindre au mieux l'espace de représentation.

B) La démarche vue comme un système de traitement de l'information

Tout processus de décision nécessite d'analyser et de synthétiser un certain nombre d'informations, que l'on peut schématiquement décomposer en 3 types : les informations existantes, les informations obtenues après recherche (par exemple par utilisation d'un modèle), et celle obtenues après traitement des informations existantes.

1) Les informations existantes

Avant de devoir définir une stratégie de pompage, le directeur technique et les chefs de poste avaient un certain nombre d'informations immédiatement disponibles et utilisables. Ce sont par exemple le type de journée vis à vis du contrat EDF, l'état de fonctionnement des équipements, qu'ils soient internes à l'usine ou concernent l'usine jumelle. Ce sont aussi, par référence aux consommations des jours précédents, des indications sur la classe de tirage du jour concerné.

2) Les informations issues d'un modèle

Les premières informations issues d'un modèles sont les prévisions de consommations journalière et horaires, qui ne sont donc connues qu'avec incertitude, ce qui impose de lier leur utilisation à celle de risque.

Le second type d'informations qui vont être pris en compte dans la décision sont les évolutions prévisionnelles issues des calculs utilisant les modèles de simulation.

Il a pu être défini deux modèles de simulation, le premier étant un modèle élémentaire, facile à mettre en oeuvre, mais ne donnant que des indications grossières. Le second produit des informations plus fines, mais au prix d'un calcul plus lourd.

L'idée que l'on veut faire apparaitre est que l'extraction de l'information a toujours un coût et que dans une démarche experte, il faut rechercher et utiliser les informations au bon moment.

3) Les informations issues d'un traitement

Très souvent les informations brutes ne sont pas utilisées telles qu'elles, soit parce qu'elles sont nombreuses, soit parce qu'elles se présentent sous des formes difficilement manipulables. La plupart des traitements consistent à les rassembler, en un ou plusieurs niveaux synthétiques, ce qui facilite leur compréhension et leur utilisation.

4) L'utilisation d'informations dans l'ordre de complexité d'obtention

Le nombre d'informations qu'un être humain peut gérer étant limité à quelques items, la manipulation d'un grand nombre d'informations nécessite de mettre en place un système qui les filtre, et n'en présente qu'un nombre limité à chaque étape du raisonnement.

Une réduction du nombre des informations disponibles à un niveau du raisonnement s'effectue soit par simple élimination des informations non

pertinentes, soit en synthésisant un certain nombre d'informations élémentaires.

Une démarche méthodologique de prise de décision peut être comprise comme un système itératif de traitement de l'information, avec deux phases principales :

- utilisation optimale des informations disponibles à un niveau donné,
- recherche ou création d'informations supplémentaires.

Il existe un certain nombre de méthodologies capables d'utiliser des heuristiques pour explorer un espace de recherche, en se basant sur un traitement des informations. Parmi celles-ci, on peut citer :

- les stratégies de gestion d'hypothèses avec un retour arrière chronologique ou sélectif, c'est à dire en remettant en cause un chemin d'exploration. C'est le choix le plus récent qui peut être remis en cause, ou on recherche la cause de l'échec avant d'ouvrir un autre noeud.
- la génération et le test d'hypothèses. C'est le cas du raisonnement du logiciel DENDRAL, qui est décrit dans le chapitre II.
- la gestion d'un raisonnement multiple par développement en parallèle d'un nombre limité de solutions partielles.
- le raisonnement par le principe du moindre engagement, c'est à dire en différant le choix jusqu'à disposer de plus amples informations pour le faire.
- la résolution hiérarchisée par affinage successif d'une solution grossièrement définie initialement, en introduisant à chaque niveau d'abstraction des contraintes et des données supplémentaires.
- l'élimination par filtrages successifs, toujours en introduisant par étapes des contraintes et données supplémentaires.

Parmi ces méthodologies, c'est sur le principe "Generate and test", où chaque niveau de filtrage s'appuie sur les informations disponibles, ou créées justes avant, que s'est orientée la réflexion de cette phase de conception.

Dans notre cas, ce sont les modèles de simulation qui vont servir à créer des informations.

Comme dans le jeu du taquin, où l'on dispose de plusieurs modèles pour la fonction d'évaluation, les différents filtrages peuvent eux-aussi être

de calcul, mais qui ne sont appliqués qu'à un nombre de plus en plus restreint de stratégies.

C) Les idées directrices de la démarche suivie

1) Optimiser l'utilisation et la recherche d'informations

La première idée directrice vient à la fois de considérations sur la démarche suivie par les experts et d'une réflexion sur l'optimisation de l'utilisation des informations, en fonction de leur accessibilité, leur pertinence, et le coût de leur recherche. Toute recherche d'information doit donc être reliée de façon pertinente à l'utilisation qui en sera faite. Une bonne démarche consiste à rechercher, à chaque étape du raisonnement, les informations minimales qui permettent une restriction maximale de l'espace de recherche.

Il en découle que la méthodologie a été basée sur la construction de cycles composés de la sélection, la recherche, et enfin le traitement d'informations dans la gestion des hypothèses de pompage. Chaque cycle correspondra de fait à un niveau de filtrage.

Un choix optionnel important est de décomposer le problème et de raisonner le plus possible à l'intérieur des sous-problèmes, et de raisonner sur des solutions globales seulement à la fin. Cela permet d'utiliser au mieux les contraintes sur les points de repère. Cette décomposition fait que sera mise en place une boucle de génération et test de solutions dans chaque sous-problème.

Dans chaque période de planification, le modèle cognitif des chefs de poste, se focalise d'emblée sur un nombre restreint de stratégies potentielles, à partir d'une utilisation des informations initialement disponibles. Le noyau de stratégies définies est ensuite examiné plus finement avant la sélection, en recherchant par un calcul sommaire un deuxième niveau d'informations.

b) Rendre explicite la démarche

La seconde idée directrice a été de pouvoir marquer et repérer, à tout moment, à quel niveau du raisonnement une stratégie a été éliminée. L'utilisation d'heuristiques était justifiée par la volonté de faire apparaître les points de décision.

La démarche va donc consister à générer et tester des solutions dans chaque sous-problème, d'abord par l'utilisation d'informations simples, puis en recherchant d'autres informations en utilisant les modèles de simulation et de prévision.

en recherchant d'autres informations en utilisant les modèles de simulation et de prévision.

Les stratégies générées peuvent ainsi être éliminées et hiérarchisées en maintenant un raisonnement cyclique, par étapes, qui permette de repérer à quel niveau du raisonnement une stratégie a été éliminée.

D) Le schéma du système

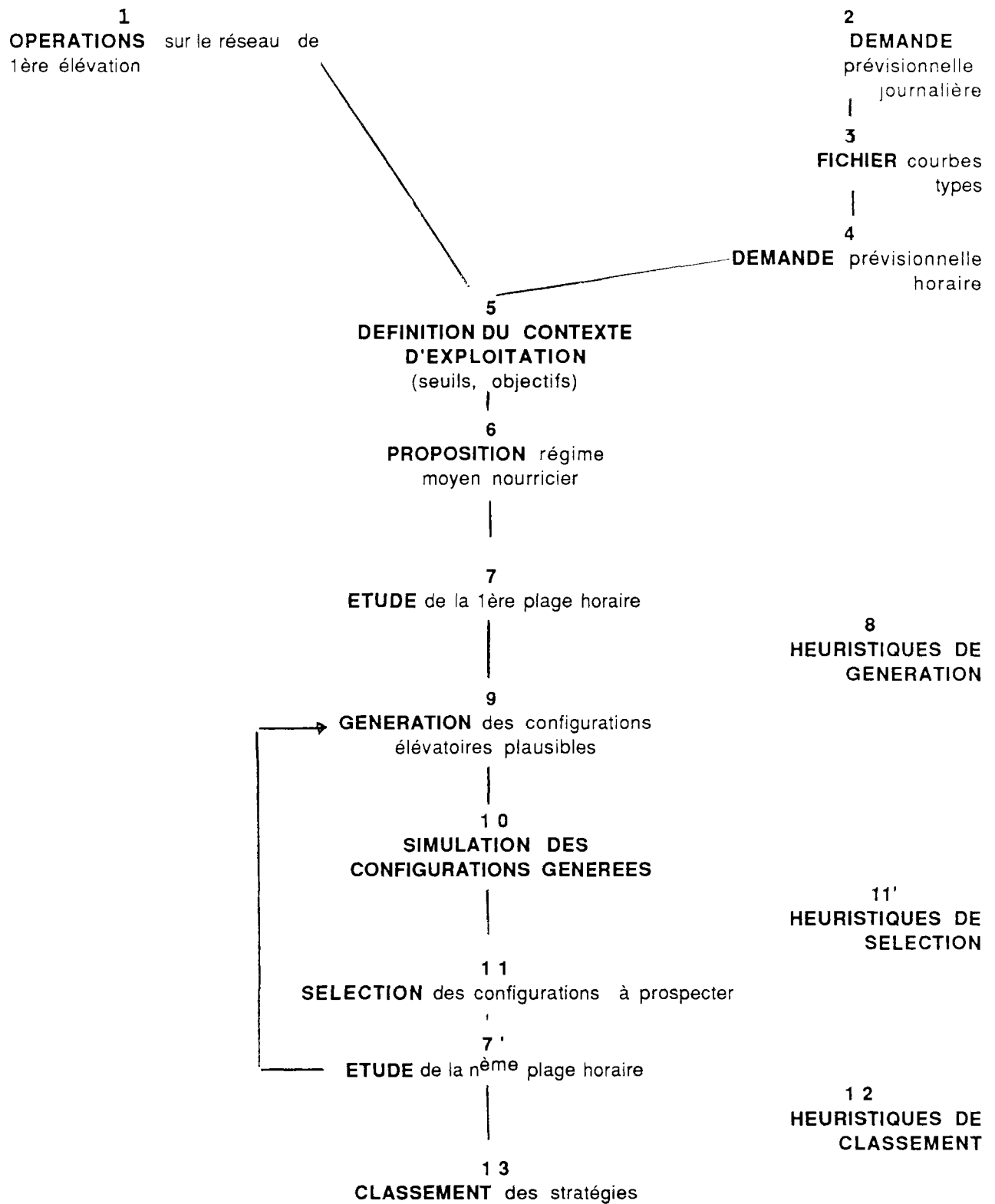
Des considérations précédentes sur la démarche suivie, il résulte que le système PILOTE a été conçu sous forme d'un ensemble de modules ayant pour fonction, soit d'utiliser au mieux un niveau d'information, soit d'en créer, soit de les traiter.

L'organigramme du raisonnement est représenté sur le schéma de la page suivante.

Les modules, qui peuvent être rassemblés en 6 groupes principaux, ont pour fonctionnalités principales :

- la création d'informations sur les prévisions de consommation,
- la définition du contexte d'exploitation de l'usine et le découpage de la journée,
- la création d'informations sur les évolutions prévisionnelles,
- la formulation d'une hypothèse de régime nourricier,
- les premiers niveaux de filtrage basés sur l'utilisation de données numériques,
- les seconds niveaux de filtrages basés sur un traitement qualitatif des informations,
- la reconstitution des stratégies complètes.

Les modèles de simulation et de prévisions, générateurs d'informations, ont été décrits au chapitre I. Les autres modules, et leur enchaînement sont décrits dans les paragraphes suivants.



II) LA DEFINITION DU CONTEXTE ET LA DECOMPOSITION DU PROBLEME EN SOUS-PROBLEMES

A) La nature de la décomposition

La reproduction de la démarche de l'expert conduit à décomposer le problème en définissant d'abord les périodes de planification sur lesquelles vont être définies les stratégies partielles.

Une journée peut être décomposée selon trois types de raisons. Le premier vient des contraintes fortes imposées par EdF sur les puissances disponibles à l'usine en fonction de la tranche horaire.

Le second type vient de l'analyse des courbes de consommation d'eau qui présentent un creux de consommation vers 14-15 heures. Ce creux est particulièrement intéressant, parce qu'il permet de faire un point relativement précis.

Le troisième type vient de la volonté de jouer sur les périodes de tarifications horaires de l'électricité, ce qui va décomposer la journée en période de stockage et de déstockage.

B) Les règles de structuration de la journée

Ces règles ont été déduites des observations relatives à la définition des contextes. On va utiliser pour définir la structure de planification les heuristiques définies dans le chapitre précédent.

Les règles consistent à déterminer le contexte d'exploitation, lequel va activer des schémas définissant :

- les périodes de planification,
- les objectifs de fin de plage,
- les fourchettes d'élimination, c'est à dire les zones de risque acceptables.

C) Exemples de règles de structuration de la journée

1) Les règles du contexte d'exploitation normale

Le contexte est défini comme normal s'il n'y a pas d'équipement indisponible (interne à l'usine, ou concernant le réservoir de première élévation ou l'usine jumelle), et si l'usine ne doit pas alimenter d'autre réseau que le sien.

Six contextes ont été définis, en croisant les informations relatives, au type de jour vis à vis du contrat EDF, à la classe du tirage, et à la nécessité de fournir ou non de l'eau au réseau Nord.

La forte demande matinale ne concerne que les jours ouvrables d'hiver, puisque, au-dessus d'une certaine demande, il sera indispensable de modifier les régimes de la nourricière pour ne pas faire déborder les réserves internes.

La première série de règles consiste donc à identifier le contexte. Elles sont du type de la suivante :

Si Demande-prévisionnelle totale > 450000
 Demande prévisionnelle (7 - 11h) < 120000
 Jour ouvrable d'hiver
 Pas de demande du réseau Nord

Alors contexte : C1 (jour ouvrable d'hiver de fort tirage)

Les contextes créés par les défaillances d'équipement ou de réservoirs n'ont pas été abordés dans le cadre de cette étude.

2) Les contextes définis dans le cadre de l'étude

C2 : jour ouvrable d'hiver de tirage moyen

C3 : jour ouvrable d'hiver de tirage fort avec demande du réseau Nord,

C4 : jour ouvrable d'hiver de tirage fort avec forte demande matinale,

C5 : jour quelconque de printemps, automne ou week-end d'hiver,
 avec un tirage moyen

C6 : jour quelconque de printemps, automne ou week-end d'hiver,
 avec un tirage moyen

3) Les règles de définition des découpages et des objectifs à partir des contextes

Une partie de ces règles définissent la structuration de la journée et fixent les objectifs associés aux heures de fin de périodes de planification (appelées plages horaires). L'autre partie de ces règles fixent l'hypothèse de régime nourricier qui sera évaluée en priorité.

Exemples :

Si Contexte est "C1" ou "C3"

Alors Découper la journée en 3 plages

Fixer les points de repère à : 5,4 m pour la plage P1
Pour "Montreuil" 4,3 m pour la plage P2
8,1 m pour la plage P3;

Et , pour R1-R2, R3,....

Si Contexte est "C1" ou "C2"

Alors Envisager un seul régime nourricier

Choisir l'option de répartition uniforme.

D) Les raisonnements à l'intérieur des périodes de planification

Tous les raisonnements basés sur le cycle "génération-simulation-sélection" ont lieu à l'intérieur des plages horaires définies lors de cette décomposition de la journée, en se servant des objectifs définis dans les règles de structuration.

A la fin de cette phase sont donc connues les périodes de planification et les objectifs de fin de plage correspondants, et une hypothèse de régime nourricier a été définie.

III) UNE PREMIERE SERIE DE FILTRAGES BASES SUR LES DONNEES INITIALES UTILISEES A PRIORI.

A) La génération du premier noyau de stratégies potentielles

L'esprit de la démarche consiste tout d'abord à définir, à l'intérieur de chaque plage horaire, le plus petit ensemble de stratégies partielles pertinentes pour la première plage horaire, et pour les plages suivantes, la meilleure stratégie.

La distinction vient du fait que la stratégie définie pour la première plage horaire va être mise en oeuvre, et qu'il importe de voir quelle suite ont toutes les stratégies jugées pertinentes sur cette première plage.

La fonction de cette phase est donc de définir un ensemble de stratégies, que l'on va essayer de faire correspondre au noyau des stratégies potentielles sur lesquelles les chefs de poste se focalisent d'emblée, en utilisant les informations structurantes issues des contextes.

Les heuristiques de génération ont été décrites dans le paragraphe précédent. Ces heuristiques sont traduites sous la forme d'un ensemble de règles qui définissent le noyau initial de stratégies envisageables. La connaissance et l'expérience de conduite d'usine conduisent à éliminer, à priori, sur la seule base de définition du contexte, tout un ensemble de stratégies.

Les règles sont reprises à partir du jeu d'heuristiques définies. Elles concernent à la fois la limitation des configurations possibles et la limitation des changements de configurations possibles dans une plage horaire.

B) La structuration de la connaissance

A chaque contexte est associé un fichier contenant les configurations de pompage envisageables. Chaque fichier a été construit en fonction des heuristiques dites "de génération", présentées dans le chapitre III, qui éliminent, à priori, sans calcul, les configurations qui n'élèvent pas assez d'eau, ou trop, ou dont la consommation électrique est trop élevée, ou encore dont le couplage entre pompes est mauvais. De plus, le nombre de changements de configurations autorisés dans la plage horaire est strictement limité.

Il est évident qu'à chaque contexte est associé un fichier particulier, puisque l'on fait intervenir la classe de tirage, et que dans certaines situations d'exploitation, des configurations aux rendements moyens devront être envisagées.

C) L'efficacité de ces premiers filtrages

Sur une période de planification d'une dizaine d'heures, on restreint ainsi l'espace de recherche de N^{10} à une centaine de stratégies.

On peut toutefois noter que les stratégies éliminées n'apparaissent nullement dans le programme, c'est à dire qu'elles n'ont pas d'"existence informatique". Ce n'est qu'à partir de ce premier filtrage que les stratégies seront indexées et donc manipulées dans les programmes.

V) UNE SECONDE SERIE DE FILTRAGES BASES SUR LES DONNEES QUANTITATIVES OBTENUES PAR UN MODELE SOMMAIRE.

A) L'utilisation des points de repère avec introduction de la notion de fourchette

1) La recherche d'informations complémentaires

Le premier filtrage s'est appuyé sur une partie des informations initialement disponibles. Quand commence la génération des stratégies partielles sur une plage horaire, un certain nombre d'informations sont encore connues, comme les volumes en réserve à Montreuil et dans les effacements. D'autre part, le modèle de prévision a donné une indication des volumes à produire pour la plage horaire et les sous-plages. Enfin, en conduite d'usine, les chefs de poste visent des niveaux de réservoir, qui sont autant de références utilisables, et qui ont été définis comme objectifs.

La reproduction de la démarche des chefs de poste à ce niveau consisterait à effectuer des calculs simplifiés pour évaluer, en première approximation les conséquences induites par chacune des stratégies du noyau.

La démarche suivie reprend cette idée en utilisant les prévisions de consommation et le modèle fonctionnel simplifié pour mettre en relation les conséquences des stratégies et les objectifs et contraintes de la gestion d'usine pour la période considérée. Ces calculs simplifiés ne prennent pas en compte ce qui se passe dans la durée de l'intervalle.

Le nombre encore très élevé de stratégies envisageables sur la période de planification fait qu'il serait très coûteux en temps de passer toutes les stratégies au crible du modèle de simulation. Et ce d'autant plus que le modèle sommaire, permet de restreindre efficacement l'espace de recherche en mettant en relation les seuils de sélection avec les simplifications dues au calcul.

2) Le choix d'une fourchette d'élimination

Si l'on comprend toujours la démarche comme l'optimisation de l'utilisation et de la recherche des informations, on peut travailler à ce niveau avec une faible précision, parce que le coût d'acquisition des informations qui permettent d'opérer la sélection n'est pas élevé.

Des points de repère ayant été définis, il est alors possible de procéder à l'élimination des stratégies, qui s'en écartent plus que l'erreur maximale faite par l'approximation du calcul. Cette première élimination est faite en utilisant une fourchette large.

Il n'empêche qu'il y a malgré tout un certain nombre de précautions à prendre, c'est à dire que les critères d'élimination des stratégies doivent être inférieurs aux erreurs de précision des calculs. Le choix de la fourchette est donc important, puisque la réduire diminue l'efficacité de la sélection, et l'augmenter fait que l'on va devoir rechercher des informations additionnelles pour un beaucoup plus grand nombre de stratégies. Il faut donc trouver le compromis entre l'efficacité et le risque de rejeter d'emblée une solution intéressante.

Le calcul des largeurs de fourchette a été effectué à partir des volumes moyens élevés pour une pompe sous la hauteur manométrique moyenne de la plage horaire. Dans les cas d'exploitation normale, les différentes hauteurs manométriques moyennes d'une période donnée ont une dispersion faible, et il a donc été possible de définir une fourchette de niveaux considérés comme acceptable.

Par exemple, si l'objectif de remplissage à Montreuil est 5 m, pour 1h du matin et la fourchette est de 1m, les limites inférieure et supérieure pour l'objectif sont respectivement :

minimum : 4 m , la raison étant que l'on ne pourra pas remplir Montreuil en partant de moins de 4 m à 1 heure du matin.

maximum : 6 m, dans ce cas de figure, il est vraisemblable que le pompage de nuit a été défavorisé.

Un calcul sommaire, effectué seulement pour les configurations, la plage et les sous-plages sélectionnées, détermine le niveau prévisionnel pour chaque stratégie partielle. Si ce niveau est compris dans la fourchette, la stratégie partielle sera examinée plus en détail, c'est-à-dire générée dans la terminologie de PILOTE, sinon elle est rejetée à ce niveau du raisonnement.

B) L'efficacité de ce second niveau de filtrage.

Ce filtrage permet d'éliminer les stratégies qui sont au-delà de la précision du calcul vis à vis des objectifs définis en fin de période. Ceci représente les trois quarts des stratégies envisageables à ce stade du raisonnement. Donc, sur la centaine de configurations retenues à priori, pour un plage horaire de 10 heures, il reste après ce filtrage, un peu plus d'une vingtaine de stratégies.

L'analyse des erreurs de planification, indiquent que celles-ci ont été dues à des erreurs sur la prévision de consommation, et non à une fourchette trop étroite.

Le modèle sommaire ne fournit aucune information sur ce qui se passe dans l'intervalle, aussi la sélection n'utilise pas les contraintes.

C) Du choix des chefs de poste à un choix informatique

Les chefs de poste, qui ne disposaient pas de moyens informatiques de calcul, choisissaient à ce niveau du raisonnement parmi leurs options possibles de configuration, l'option moyenne.

Mais l'ensemble de configurations envisagées par les chefs de poste, à ce niveau de leur démarche, dépassait rarement 5. Continuer un filtrage nécessite l'acquisition d'informations complémentaires, et en conséquence, des moyens de calcul. L'élément qui peut apporter des informations est le modèle fin de simulation.

Cette recherche et utilisation d'informations complémentaires, basée sur un modèle fin, va être décrite dans le chapitre suivant.

D) Une première étape vers l'apprentissage

A ce niveau, il est possible de faire un suivi indiquant le nombre des erreurs par des calculs à postériori, en voyant la part due aux incertitudes sur les prévisions de consommation. On peut ainsi espérer procéder à un ajustement automatique de la largeur de la fourchette.

V) UNE TROISIEME SERIE DE FILTRAGES BASES SUR LES DONNEES QUANTITATIVES OBTENUES PAR LE MODELE DE SIMULATION.

A) L'introduction d'un second niveau dans le raisonnement

Au premier calcul simplifié va succéder une simulation des évolutions des grandeurs hydrauliques et énergétiques dans l'usine, en utilisant le modèle fonctionnel et le modèle prévisionnel, tous deux décrits au chapitre I. La simulation est faite avec un pas de temps horaire et sur la durée de la période de planification.

Les résultats beaucoup plus fins obtenus pour chaque stratégie permettent d'affiner la sélection, par la création d'un grand nombre d'informations relatives à chaque stratégie.

B) Les informations apportées par le modèle de simulation

1) Une description quantitative sur le déroulement prévisionnel de la plage horaire.

Cette fois, le modèle de simulation fournit des indications sur les évolutions prévisionnelles sur toute la durée de la plage horaire.

Il est donc possible d'utiliser les contraintes, exprimées comme des seuils relatifs à des niveaux ou à des puissances électriques, pour éliminer un certain nombre de stratégies. La simulation travaille comme un filtre.

2) Une meilleure précision sur les états prévisionnels de fin de plages

Le modèle de simulation donne des informations sur les niveaux en fin de plage qui sont plus précises que celles du modèle sommaire. Il est donc alors possible de restreindre la fourchette d'écart à l'objectif à partir duquel une stratégie est éliminée.

B) L'efficacité de ce second niveau de sélection numérique

L'utilisation des contraintes et la restriction de la fourchette d'acceptabilité par rapport à l'objectif de fin de plage permet de restreindre une nouvelle fois l'espace de recherche à sa moitié, ce qui réduit l'ensemble des stratégies envisageables, pour une plage horaire de 10 heures, à environ une dizaine.

C) Le traitement des informations issues de la simulation

La simulation d'une stratégies sur une plage horaire apporte un grand nombre d'informations, qui n'ont pas toutes le même degré de pertinence et qu'il est très difficile de traiter de façon brute. En effet, une simulation décrit les évolutions de trois réservoirs, de plusieurs puissances électriques, et pressions. Une simulation de stratégie sur une durée de 8 heures produit un fichier caractérisé par une dizaine de champs, soit 80 données.

Pour traiter efficacement ces informations, pour continuer le processus d'élimination des stratégies, il est nécessaire de les regrouper dans un niveau synthétique qui soit gérable et compréhensible. C'est la raison pour laquelle on va utiliser un raisonnement qualitatif, en prenant appui sur les classes de qualification des critères et sur leur hiérarchie, telles qu'elles résultent de l'analyse des réponses aux questionnaires décrits précédemment.

V) LA SELECTION D'UNE STRATEGIE PAR UN RAISONNEMENT QUALITATIF

A) La formulation du problème

A ce niveau de la démarche suivie, les stratégies sont passées par 3 filtres :

- génération d'un noyau de stratégies potentielles, définies à priori, par la connaissance des éléments principaux du contexte, avec en conséquence, l'élimination de toutes les autres,
- premier filtrage basé sur les objectifs définis et sur un calcul simplifié,
- second filtrage, basé sur les mêmes objectifs définis, mais suite à une simulation fine.

A ce stade du processus d'élimination des stratégies, il reste une dizaine de stratégies par plage horaire, soit donc une centaine ou un millier encore envisageables selon que la journée a été divisée en 2 ou 3 périodes.

Or la phase de simulation a apporté un grand nombre d'informations sur les évolutions prévisionnelles, informations qu'il apparaît très intéressant d'utiliser avant de connecter les stratégies sur les différentes plages.

L'idée est donc d'utiliser des informations en raisonnant qualitativement pour continuer à restreindre l'espace de recherche à l'intérieur des plages horaires avant de recomposer le problème.

Mais comme cela a été dit précédemment, il est impossible de faire efficacement des choix en manipulant un aussi grand nombre d'informations que celui apporté par chaque simulation.

Prolonger la démarche nécessite deux opérations successives. La première consiste à extraire les informations pertinentes de l'ensemble des données apportées par la simulation. La seconde consiste à définir une méthode de traitement de ce noyau d'informations pertinentes.

B) L'extraction des informations pertinentes

Le noyau à extraire a été défini à priori comme le même que celui utilisé par les chefs de poste lorsqu'ils analysent les qualités d'une stratégie de pompage, c'est à dire :

- la plus petite différence entre le niveau du réservoir de première élévation et le seuil défini,
- la plus petite différence entre la puissance appelée et la puissance limite définie dans le contrat,
- la distance à l'objectif de fin de plage horaire,
- la pression maximale de refoulement atteinte,
- le coût de la stratégie,
- le nombre de changements de configurations dans la plage horaire.

Cette phase, basée sur l'expertise recueillie et décrite précédemment, a servi à extraire de l'ensemble des données, 7 éléments numériques pertinents pour les comparaisons ultérieures de stratégies.

Mais les experts manipulent rarement des valeurs numériques, car leur combinaison fait perdre tout caractère explicite au raisonnement. Si l'on veut pouvoir garder une trace des raisonnements, et plus précisément dans notre cas, les causes et phases d'élimination des stratégies, il est indispensable, comme le remarque d'ailleurs VOGEL [VOGEL 88], de remplacer les grandeurs variant continuellement par des catégories discrètes.

C) Le regroupement des critères en niveaux de qualification

En se servant des réponses aux questionnaires faites par les chefs de poste lors du recueil de l'expertise, la phase de comparaison des stratégies se fera sur la base d'un tableau de 7 critères, dont les valeurs "Bon, moyen, muavais", ont été extraites des tableaux de données résultant de la simulation fine.

Ces critères se présentent dans un tableau comme celui ci-dessous, et caractérisent toutes les stratégies encore sélectionnées à ce niveau :

STRAT. N°	SECU Eau	* SECU * EdF	* HMT *	* DIS * ojb	* Regi * Nour	* Chgt * Elev	* COUT *
1	*	*	*		*	*	*
2	*	*	*	*	*	*	*

C'est donc à partir des informations contenues dans un tel tableau que le module de classement de PILOTE doit déterminer la meilleure stratégie à suivre.

D) La présentation de deux méthodologies envisageables

La recherche bibliographique, complétée par une réflexion d'adaptation, a abouti à définir deux méthodologies qui pouvaient être utilisées dans le cadre du raisonnement :

- l'établissement d'un coût global,
- le classement par déqualification.

Ces deux voies ont été successivement explorées.

1) Choix à partir du calcul d'un coût global :

a) L'idée :

Cette méthode consiste à traiter de façon particulière le coût, gardé comme une valeur numérique, et à le multiplier par un pourcentage dépendant de la déqualification de la stratégie.

b) Les éléments constitutifs du surcoût :

Trois types de critères (très important, important et secondaire), et quatre classes de déqualifications (totale conduisant à l'élimination, forte, faible et nulle) ont été définis. Ces classes correspondent bien sûr aux catégories "Mauvais, moyen et bon", définies précédemment. La méthode consiste ensuite, à définir cas par cas, un pourcentage de surcoût, tel qu'il est représenté dans le tableau suivant.

IMPORTANCE	* TRES IMP	* IMPORTANT	* SECONDAIRE	*
DEQUALIFICATION	*	*	*	*

TOTALE	* ∞	* ∞	* ∞	*

FORTE	* X_{ti}	* X_i	* X_s	*

FAIBLE	* Y_{ti}	* Y_i	* Y_s	*

NULLE	* 0	* 0	* 0	*

Ainsi X_{ti} représente le pourcentage de surcoût affecté à un critère défini comme très important, et ayant une déqualification forte.

Il est évident qu'à une déqualification nulle correspondra un surcoût nul.

c) Calcul du coût global:

On passe en revue l'ensemble des caractéristiques de la stratégie en appliquant pour chaque critère le pourcentage de surcoût de la catégorie lui correspondant. Tous ces pourcentages sont ensuite additionnés. Le coût global est le produit du coût énergétique par la somme des surcoûts des critères.

$$C_g = C \cdot (\sum \text{surcoûts})$$

La stratégie choisie sera celle dont le coût global C_g est le plus faible.

2) Déqualification par les poids et les utilités :

a) L'idée :

Cette méthode consiste à définir pour chacun des sept critères d'une part un poids lié à l'importance qu'on lui accorde, et d'autre part une échelle d'utilité donnant un coefficient à chaque classe de déqualification.

Une combinaison linéaire des poids et des coefficients permet de définir pour chaque stratégie un indice de déqualification.

Cette méthode implique de considérer le coût énergétique comme un critère comme et les autres et donc de définir ses classes de déqualification.

b) Poids et échelle d'utilité :

Les poids des critères seront définis à partir de leurs importances absolues et relatives telles qu'elles ont été définies dans les réponses aux questionnaires.

La notion d'échelle utilité a pour fonction d'accorder plus ou moins d'importance à une déqualification donnée pour un critère particulier.

Dans l'exemple ci-dessous, les stratégies ayant une déqualification faible vis à vis du critère sécurité en eau seront fortement pénalisées, tandis que celles ayant une déqualification faible vis à vis du critère sensibilité le seront beaucoup moins.

CRITERE	* SECU-Eau	* DIST-obj	* Sensibilité	*
DEQUALIFICATION	*	*	*	*
TOTALE	* ∞	* ∞	* ∞	*
FORTE	* 9	* 8	* 9	*
FAIBLE	* 7	* 4	* 3	*
NULLE	* 0	* 0	* 0	*

c) Indice global de déqualification et choix d'une stratégie:

L'indice global de déqualification d'une stratégie est égal à la somme sur l'ensemble des critères, des produits des poids p_i par les utilités u_i .

La stratégie retenue est celle dont la somme $I_d = \sum(p_i \cdot u_i)$ est la plus faible.

E) Les limites de ce type d'approche

1) Un raisonnement étranger aux chefs de poste

Les méthodes multi-attribut décrites précédemment nécessitent la translation de la connaissance des preneurs de décision en fonction analytique qu'il leur est ensuite très difficile de gérer pour vérifier, justifier et expliquer le modèle lui-même, et expliquer comment les résultats ont été obtenus. Les essais de remplissage de ce type de grilles par les chefs de poste se sont vite révélés négatifs. Le raisonnement leur demandait de manipuler des notions étrangères à leur expertise, ce qui ne pouvait donner des résultats pertinents. Mais l'analyse de cet échec a fait ressortir le point essentiel auquel tenaient les experts : l'explicabilité de la décision.

2) L'explicabilité de la décision

Il est apparu lors des entretiens que les décideurs souhaitaient garder explicite l'ensemble des démarches qui justifie la sélection effectuée, car comme il n'existe à priori aucune décision optimale, il est important de retrouver la trace du choix, et de pouvoir ensuite la ressortir pour l'analyser à postériori.

F) La méthodologie retenue : la sélection lexicographique par critères

1) Le principe de la méthode

Les informations disponibles pour définir un mécanisme de sélection étaient la hiérarchie consensuelle des critères, et leurs classes de qualification. Dans un premier temps, la méthode la plus simple et la plus proche du raisonnement des chefs de poste a été une sélection lexicographique par critère, telle qu'elle est présentée par Voogd [VOOGD 83], pour arbitrer des choix d'aménagements territoriaux.

Cette méthode avait l'avantage de rendre compréhensible et d'explicable le mécanisme de décision, pour sélectionner les stratégies partielles dans une période de planification. Cette méthode prend d'abord en compte le critère de décision classé premier dans la hiérarchie : la sécurité d'approvisionnement en eau. Si une stratégie surclasse toutes les autres sur ce critère, elle est retenue. S'il y a égalité parmi les stratégies restantes, la décision est basée sur le second critère de la hiérarchie : la sécurité vis à vis du contrat EDF. Le processus continue jusqu'à ce qu'il n'y ait plus d'égalités.

2) L'algorithme de décision

L'algorithme retenu dans la maquette est basé sur une utilisation stricte de la hiérarchie décrite précédemment. Une telle procédure de décision est dite lexicographique.

Pour chacun de ces critères, les valeurs numériques ont été regroupées en 3 classes correspondant aux 3 niveaux de qualifications "bon, moyen, mauvais". L'algorithme est le suivant :

* La première étape est le comptage du nombre n de stratégies de l'ensemble à sélectionner, ayant "bon" comme qualification pour le premier critère. Trois cas peuvent se présenter, représentés sur le schéma de la page suivante :

- a) $n = 1$ la seule stratégie ayant la qualification "bon" est alors sélectionnée
- b) $n = 0$ Le comptage reprend, mais pour la qualification "moyen".
 - b.1) $n' = 0$. Le 1er critère n'est pas discriminant.
Passage au critère suivant
 - b.2) $n' = 1$ La seule ayant la qualification "moyen" est alors sélectionnée.
 - b.3) $n' > 1$ Restriction de l'ensemble à sélectionner aux seules stratégies ayant cette qualification
- c) $n > 1$ Le critère placé en seconde position dans la hiérarchie est alors pris en compte, et traité de manière similaire

La sélection est terminée quand l'ensemble des stratégies est réduit à une unité.

VII) LA SELECTION D'UNE STRATEGIE COMPLETE

A) La reconstitution des stratégies complètes

Dans la première plage horaire ont été définies n stratégies partielles. Toutes les stratégies potentielles après les phases de génération et de simulation constituent autant de stratégies parent pour les plages horaires suivantes de la journée.

A partir de la seconde plage, le cycle de génération, simulation, et sélection, ne retient qu'au plus une stratégie "fille", par stratégie "parent" de la première période. La reconstitution des stratégies complètes est donc une simple opération de connection des stratégies partielles, de la première à la dernière période de planification.

Le nombre de stratégies complètes est donc égal ou inférieur au nombre de parents définis dans la première période. Il y a égalité quand, tous les parents sont au moins une stratégie partielle non éliminée par les contraintes ou par une distance trop grande à l'objectif.

B) Les cas de remise en cause de l'hypothèse de régime nourricier

Si aucune stratégie complète n'a pu être définie, il existe un certain nombre de cas où c'est le régime nourricier initialement défini, qui est la cause des éliminations.

Plusieurs de ces cas ont été répertoriés, comme le franchissement des limites des réservoirs internes, ou la non satisfaction de la demande en cours de journée, mais avec une réserve de puissance électrique.

Ces situations conduisent à une nouvelle formulation du régime nourricier, prenant en compte la cause de l'échec. Cette nouvelle hypothèse propose dans tous les cas, au moins un changement de régime, qui dépend des heures où l'échec des stratégies ont été constatées.

La formulation d'une nouvelle hypothèse impose la réinitialisation de l'ensemble de la boucle "génération-simulation-sélection", sur les plages horaires de la journée.

C) Le classement parmi les stratégies complètes

La base de la méthode utilisée est la même que pour les stratégies partielles. Il est cependant nécessaire de reprendre complètement la phase de qualification des stratégies, avant d'appliquer l'algorithme de décision lexicographique.

Comme le nombre de stratégies complètes sélectionnées est faible (rarement plus de 5), il a été décidé de présenter l'ensemble de ces stratégies, dans l'ordre de leur classement, en visualisant leurs principales caractéristiques.

CH V : IMPLEMENTATION DU LOGICIEL PILOTE

I) L'IMPLEMENTATION DE LA METHODOLOGIE

A) Le problème posé

La méthodologie de résolution ayant été définie et présentée dans le chapitre précédent, il importe maintenant de définir les outils et leurs caractéristiques nécessaires pour construire une maquette.

Un piège toutefois était à éviter. On ne construit pas un système expert en utilisant simplement un langage déclaratif ou un logiciel générateur de système expert. C'est la démarche suivie et non l'outil qui justifiera le terme.

Cependant, le choix des outils de développement s'est posé concrètement dans le cadre du déroulement de la thèse, et la sélection d'un outil particulier dans un marché qui en propose de très nombreux, demande une bonne compréhension des fonctionnalités et potentialités des logiciels proposés sur le marché.

Aussi, dans l'esprit de cette thèse qui consiste à tester les méthodes et outils de développement de l'intelligence artificielle dans le domaine de la gestion hydraulique, il est apparu important de détailler les alternatives possibles, et d'indiquer les considérations qui ont prévalu lors des choix.

De nombreux outils informatiques pour l'aide au développement de systèmes-experts existent aujourd'hui sur le marché. D'une manière générale, ces outils comportent deux classes de fonctionnalités correspondant respectivement au constructeur des bases de connaissances, et à son utilisateur final.

Selon les outils, les classes sont plus ou moins développées, et même la seconde totalement vide pour les versions standard des langages de programmation symbolique, comme Lisp ou Prolog.

Les outils d'aide au développement ne sont pas universels, et chacun d'eux utilise une représentation particulière des connaissances, et des mécanismes spécifiques de contrôle du raisonnement.

Présenter tous les logiciels présents sur le marché serait long et fastidieux, et ne correspond pas à l'objectif de ce chapitre de thèse, qui est de donner des points de repère. D'autre part, les caractéristiques de ces outils évoluent très vite et tout choix n'a de sens qu'à une date donnée. La description de deux langages déclaratifs, Lisp et Prolog, et de quatre logiciels

représentatifs, donnée en annexe, ne correspond certainement plus à la réalité de fin 1989. Les quatre logiciels décrits sont :

- GURU, qui peut être défini comme un générateur de système-expert combiné aux outils de bureautique traditionnels,
- NEXPERT-Objet qui est un générateur basé sur la manipulation d'objets.
- EXPERT-KIT qui est un générateur basé sur la logique d'ordre 1.
- SNARK qui est un générateur basé sur la logique d'ordre 1 pouvant définir des relations d'ordre 2, mais nécessitant une structuration de la connaissance sous forme de relations binaires.

Ensuite, seront présentées les considérations qui expliquent pourquoi un langage déclaratif n'a pas été retenu, puis le premier choix, ses limites et le choix qui a été fait ensuite.

B) Les critères de choix d'outils de développement

Outre le prix de l'achat initial, l'appropriation d'un outil informatique demande un investissement en temps qui est loin d'être négligeable. Si la finalité du projet est de se familiariser avec la notion de système-expert, le choix du logiciel n'est pas primordial, par contre, dans le cas où une réalisation doit suivre, la sélection d'un logiciel doit faire intervenir plusieurs types de critères, comme :

- le formalisme de représentation des connaissances,
- les stratégies de résolution potentielles,
- les possibilités de suivre et de visualiser les raisonnements
- le degré de finition de l'outil, et son avenir en termes d'évolution et de maintenance,
- le degré de difficulté de formation à l'outil,
- les supports de développement, vus en termes d'abondance et clarté de la documentation et en possibilités d'assistance par la société distributrice du logiciel.

1) Choisir un langage ou un générateur ?

Sans entrer dans le détail, il est évident qu'un langage permet une meilleure adéquation entre ce qui est spécifié et réalisé, mais au prix d'un développement informatique beaucoup plus lourd.

Partir de LISP, et dans une moindre mesure de PROLOG, aurait dérivé du travail de thèse, car obligé à un développement informatique important, ce qui n'était pas possible sans une aide extérieure pour la programmation.

Cette considération n'est pas la plus importante. Un examen même sommaire de la démarche méthodologique retenue fait apparaître un découpage modulaire. Or, si certains des modules se prêtent bien à l'utilisation d'un langage déclaratif, il est évident que pour d'autres, ce choix aurait été peu judicieux. La simulation, par exemple, nécessite des calculs numériques nombreux auxquels un langage déclaratif se prête mal.

Aussi, il est vite apparu que certains modules devaient être traités à partir d'un langage procédural, adapté au calcul. Ceci conduisait à séparer la connaissance, des calculs répétitifs, et, en conséquence, à prendre en compte un nouvel élément : la communication entre les modules dits experts et les modules procéduraux.

Cet élément éliminait pratiquement le langage LISP, et le générateur Expert-Kit, qui n'était interfaçable qu'avec Le_Lisp.

2/ Le formalisme de représentation des connaissances

Cet aspect inclut le degré de généralité nécessaire, c'est-à-dire la nécessité de raisonner sur des collections d'objets, avec des relations entre objets, ou sur les relations elles-mêmes.

Le second aspect prend en compte la clarté du raisonnement, c'est-à-dire la représentation qui est naturellement la plus proche de celle du domaine considéré.

La méthodologie décrite précédemment fait intervenir les notions d'objets et de classes d'objets. Ainsi les stratégies constituent des objets caractérisés par un ensemble de propriétés comme leurs qualifications par rapport aux différents critères définis.

La possibilité de regrouper les objets dans des classes et de raisonner sur des ensembles constituait donc une potentialité importante d'un outil de développement. C'est en ce sens que Nexpert et Snark présentaient un intérêt par rapport à PROLOG et GURU.

3/ Les stratégies de résolution potentielles de l'outil

La reproduction d'un raisonnement d'expert peut être le mieux approchée par un chaînage avant ou arrière. Dans certains problèmes, la prise en compte de variables variant en fonction du temps, demande que le générateur puisse raisonner en logique non monotone. Dans le système, le temps n'intervient pas, et la seule remise en cause d'une hypothèse est celle du régime nourricier, mais celle-ci implique une réinitialisation totale de toutes les variables.

La démarche d'utilisation des différents filtres a conduit à préférer un raisonnement en chaînage arrière où chaque noeud de l'arbre apparaît comme un niveau de regroupement ou de filtrage. Dans le cas où l'on veut se focaliser vers la sélection de la meilleure stratégie, le chaînage arrière paraît plus approprié, et ceci constituait un premier élément en faveur de Nexpert au détriment de SNARK.

4) La visualisation du raisonnement

Un élément important du choix était la possibilité de pouvoir suivre les différentes étapes de la sélection des stratégies, pour repérer quand une stratégie donnée était éliminée.

La possibilité de suivre, en temps réel, un raisonnement sur l'arbre des règles a donc été retenue comme une potentialité importante de Nexpert, correspondant à la volonté de rendre explicite la démarche.

5/ Le degré de convivialité de l'outil

Ce critère comprend, d'une part les facilités de développement et de mise au point, destinées au cognitif, et d'autre part, le degré de convivialité attendu par ses utilisateurs.

Pour le développement, c'est Nexpert qui présentait le meilleur environnement.

La finition d'un système ne se mesure pas qu'en termes de fonctionnalités de présentation et/ou de graphisme, mais aussi en termes de rapidité d'exécution. Cet élément est toutefois très difficile à évaluer avant d'avoir travaillé concrètement avec l'outil.

6/ Les supports de développement :

L'investissement en temps d'apprentissage à l'outil peut être considérablement réduit (ou allongé) en fonction de la clarté des documents fournis avec le logiciel et de la disponibilité de la société qui commercialise le logiciel. La plupart des sociétés proposent une formation de 2 ou 3 jours, mais celle-ci se révèle généralement insuffisante quand on est aux prises avec des problèmes concrets, comme l'interfaçage avec d'autres logiciels.

Même si cela n'a pas constitué un élément déterminant de choix, le fait de pouvoir travailler en collaboration avec EUREQUIP, société distributrice de NEXPERT, a pesé dans le choix de ce logiciel.

7) Les outils informatiques du logiciel PILOTE :

Quelques autres considérations ont conduit au choix de NEXPERT-Objet comme générateur de système expert.

Nexpert permet d'assurer la portabilité sur n'importe quel type de machine. En effet les versions IBM-AT, MAC et VAX de NEXPERT sont développées en parallèle et donc le choix de ce logiciel ne diminue en rien l'étendue des choix ultérieurs.

Nexpert offre un accès facile aux logiciels standard tels que les gestionnaires de bases de données (DBASE3), les tableurs (LOTUS) ou les outils de développement graphique.

C) La configuration retenue

1) La première option

Le couple initial d'outils de développement était formé de Nexpert-Objet, avec le langage C. Le programme de simulation a d'abord été écrit en C, pour faciliter l'interfaçage avec Nexpert-Objet. L'avantage supposé de ce choix était de permettre un échange direct en mémoire vive des valeurs des variables entre les bases de connaissances et les programmes de calcul, sans passer par un fichier.

Une version de PILOTE a donc été développée avec ces outils et présentée au Congrès EAU-2000 à Nice, en septembre 87, sur le stand d'Euréquip.

Pilote a été développé avec la version IBM-AT de Nexpert-Objet, ce qui n'a pas manqué de poser un certain nombre de difficultés liées à la limite des 640 Kilo-octets gérés par le DOS. Cette version, essentiellement destinée à la démonstration s'est ensuite révélée très difficile à gérer pour 2

raisons. La première était la lenteur d'exécution du programme, car Nexpert, sous Windows, occupe une part importante de la mémoire vive disponible sur les PC, et laisse très peu de place aux programmes et bases de connaissances. La seconde raison venait des difficultés à maîtriser les échanges de données avec le langage C sous Windows.

Comme il est vite apparu que l'intérêt d'échanger des variables, sans passer par l'intermédiaire d'un fichier, n'était pas primordial pour l'illustration de la méthodologie, une seconde option de développement a été définie.

2) La seconde option

Cette option consiste à séparer, dans le déroulement d'une session, les bases de connaissances des programmes de calcul.

Le rôle de la base est alors de définir les paramètres d'un programme de génération, simulation, et sélection des stratégies, qui applique automatiquement les indications définies dans les jeux de règles.

Cette séparation rend plus efficace les programmes de calcul, et facilite la maintenance de la connaissance dans la mesure où celle-ci est concentrée dans des bases.

C'est cette organisation qui a été présentée, le 10 mai 1988, à l'usine de Neuilly/Marne, en présence des responsables et utilisateurs concernés de la Générale des Eaux, du Syndicat des Eaux de l'Ile De France, et du CERGRENE. Cette version, décrite dans le chapitre suivant, a été évaluée sur le site, et les résultats de cette phase sont contenus dans le chapitre 6.

II) LA STRUCTURE DU SYSTEME

A) Architecture de la maquette PILOTE

1) Une décomposition en modules

Comme cela a été décrit dans le chapitre précédent décrivant la méthodologie suivie, le projet PILOTE a été conçu sous la forme d'un ensemble de modules.

La maquette informatique a donc été construite sous la forme d'un ensemble de bases de connaissances et de programmes reprenant la démarche méthodologique.

2) Les bases de connaissances et les programmes

L'enchaînement du raisonnement est représenté sur le diagramme de la page suivante.

Le programme PREVISIO correspond au module d'élaboration de la demande prévisionnelle en eau.

Les bases CONTEXTE.KB ET REGINOUE.KB correspondent respectivement à la définition du contexte d'exploitation et du régime nourricier.

Le programme PILOTE est décomposé en plusieurs modules dont plusieurs d'entre-eux forment une boucle de génération, simulation, et sélection des stratégies sur une plage horaire.

Ces modules sont complétés par une procédure de regroupement des plages horaires sur la journée complète, et par une dernière procédure de classement.

3) Le passage des valeurs entre bases de connaissances et programmes

Une tentative de passage direct des valeurs a été faite, mais devant les difficultés informatiques, ce type de transfert a été abandonné, et remplacé par un passage par l'intermédiaire de fichiers.

4) Le schéma du passage

Le diagramme ci-dessous explique la répartition des rôles entre les différentes bases de connaissances et les programmes. Cette répartition a été conçue de façon à ce que les bases de connaissances contiennent la part explicite du raisonnement et les programmes la part répétitive des calculs.

Programmes

Echanges

Bases de connaissances NEXPERT

PREVISIO

Lecture des consommations des jours précédents,
Identification
Intégration dans la demande des variations de niveau des réservoirs,
Calcul des 24 valeurs de la demande prévisionnelle totale,
Affichage de ces valeurs sous forme d'un graphe,
Sauvegarde de ces valeurs dans un fichier,

CONTEXTE.KB

Détermination de la situation d'exploitation,
Définition des plages horaires des boucles génération-simulation,
Choix des seuils et détermination des distances maximales aux objectifs,
Définition des configurations interdites à priori
Définition de l'ordre et du mode d'agrégation des critères.

REGINO.U.KB

Détermination d'un régime minimisant les changements,

BOUCLE DE CALCUL sur les N plages horaires

GENERA

Génération de l'ensemble des stratégies potentielles sur la plage horaire considérée

SIMULA

Simulation de l'ensemble des stratégies potentielles,

Suivi de l'évolution des niveaux de réservoirs, puissance appelée, HMT,

Sauvegarde de ces valeurs dans un fichier.

Elimination des stratégies potentielles dépassant les seuils ou la distance maximale à l'objectif,

Sauvegarde des caractéristiques de la situation prévisionnelle engendrée par une stratégie potentielle sélectionnée dans un fichier.

SELECTIO

Pour les plages autres que la 1ère,
Sélection d'une configuration parmi celles qui sont issues de la même stratégie de 1ère plage horaire,

REGROUPEMENT

des stratégies partielles en stratégies complètes

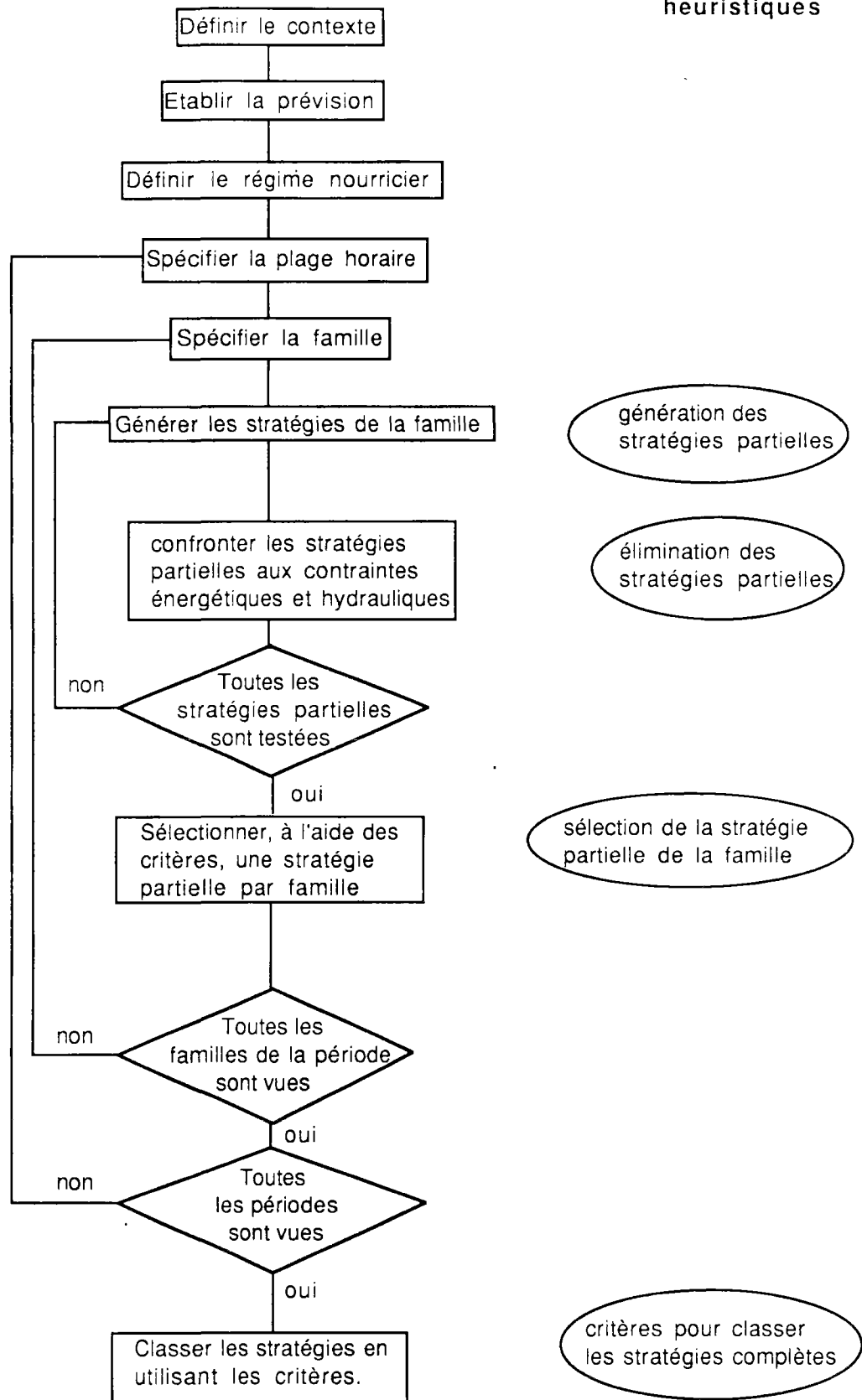
CLASSEMENT : de l'ensemble des stratégies complètes en fonction de critères donnant un ordre dans l'évaluation des suivis de l'évolution des niveaux de réservoirs, puissance appelée, HMT.

Présentation graphique de la stratégie complète proposée par le système.

5) L'organigramme du système (voir page suivante)

Modèle statistique
de consommation

Modèle de
simulation
hydraulique



Structure de PILOTE

B/ Description du programme PREVISION

1/ Fonctionnalités du programme

Les fonctionnalités de ce programme sont :

- Calcul de la consommation prévisionnelle totale du jour étudié, à partir d'une équation de régression linéaire prenant en compte le type de jour de la semaine, la consommation de la veille et la température prévisionnelle.
- Calcul des 24 valeurs de la consommation prévisionnelle journalière et intégration de ces valeurs dans un fichier ("PRE+date") pour un contrôle ultérieur.
- Calcul des 24 valeurs de la demande prévisionnelle journalière en prenant en compte la consommation, les mouvements prévus sur les réservoirs et les intercommunications.
- Affichage du graphe de la demande à l'écran.
- Sauvegarde des 24 valeurs validées dans un fichier.

2/ Les entrées :

a) *Les variables :*

- la date complète (jour-mois)
- la température maximale prévisionnelle du jour,
- les débits prévisionnels aux intercommunications,
- les mouvements de vidange ou remplissage sur les réservoirs.

b) *Les fichiers :*

- le fichier "PRE+date" contenant les consommations horaires, niveaux des réservoirs de 1ère élévation et débits aux intercommunications de la veille. Ce fichier est construit par programme de SOPHIA et est transféré par disquette dans PILOTE.
- Le fichier PROFIL contenant l'ensemble des profils de consommations quotidiennes, indexé par jour de la semaine et saison.

3) Les sorties (fichiers) :

- Le fichier "PRE+date", contenant les valeurs prévisionnelles de consommation horaire.
- Le fichier DEMANDE, contenant les valeurs de la demande totale prévisionnelle horaire (intégrant intercoms et réservoirs).

C/ Description de la base de connaissances "CONTEXTES- CRITERES"

1) Fonctionnalités

- Définition du contexte d'exploitation de l'usine,
- Définition des plages horaires de génération, simulation et sélection des stratégies.
- Définition des objectifs de fin de plage
- Affectation d'une valeur aux seuils (réservoir, puissance électrique appelée) en fonction de la situation de l'usine.
- Affectation d'une distance maximale aux objectifs définis par la situation d'exploitation de l'usine.
- Définition des classes de qualification pour chaque critère,
- Définition du mode d'agrégation lors de la prise en compte des critères

2) Organisation de la base

a) Les objets de la base

- Les plages horaires définies par les heures de commencement et de fin de plage.
- L'ensemble des seuils des réservoirs et d'approche du contrat électrique.
- Les objectifs de remplissage des réservoirs.
- Les distances à ces objectifs conduisant à une élimination des stratégies.
- Les critères définis par une classe de qualification.

b) Le sens des règles

L'hypothèse terminale est ici "SITUATION D'EXPLOITATION DETERMINEE".

Les règles qui y conduisent sont :

- SITUATION NORMALE,
- DEFINITION HEURES DES BILANS PREVISIONNELS,
- DEFINITION OBJECTIFS,
- DEFINITION QUALIFICATION ET AGREGATION DES CRITERES

Le système vérifie initialement que la situation est répertoriée et qu'un raisonnement connu peut être utilisé. Si ce n'est pas le cas, le système rend la main. Ultérieurement il est prévu d'intégrer d'autres modules indiquant le raisonnement à suivre dans les situations d'exploitation qui ont simplement été identifiées aujourd'hui (annonce d'une pollution, secours, réservoir d'effacement indisponible).

Le système définit ensuite les limites des plages horaires, les seuils d'élimination des stratégies.

Enfin un groupe de règles fixe les objectifs et les distances acceptables par rapport à ces objectifs.

Enfin un dernier groupe de règles définit les classes de qualification des critères.

3) Entrées et sorties de la base :

a) Entrées :

- le type de jour : jour ouvrable d'hiver, de printemps, week-end,
- la situation vis à vis du contrat EDF (normale, EJP...).

b) Sorties :

- fichier contenant les valeurs numériques des seuils et objectifs de fin de plage, et le mode de classement des critères.

F) La base de définition de l'hypothèse du régime nourricier

1) Fonctionnalités

Détermination du régime nourricier horaire en se basant sur la minimisation du nombre de changements.

2) Entrées

- Le fichier DEMANDE indiquant la demande prévisionnelle horaire.

3) Sorties

- Le fichier DEBITNOU indiquant les 24 valeurs du régime nourricier.

4) Organisation de la base

a) Le sens des règles :

L'hypothèse terminale est FIN OPTIMISATION CHANGEMENT, et les règles qui y conduisent sont :

- TEST 1 CHANGEMENT,
- TEST 2 CHANGEMENTS,
- TEST 4 CHANGEMENTS.

Dans cette base, des tests sont faits dans l'ordre indiqué ci-dessus pour minimiser le nombre de changements de régime nourricier. Les contraintes contenues dans ces tests sont détaillées dans le paragraphe suivant.

Les paragraphes suivants présentent en exemple, les tests successifs pour rechercher une hypothèse de régime nourricier, pendant la période de tarification EDF d'hiver et pour un jour ouvrable.

Soient :

- WEFFAC : Volume d'eau dans le réservoir d'effacement
- WEFFACMAX : Volume maximal d'eau dans ce réservoir
- WELEV : Volume d'eau dans le réservoir de 1ère élévation
- WELEVMAX : Volume maximal d'eau dans ce réservoir.

- NOUMOYth : volume horaire théorique moyen que devrait fournir la nourricière pour répondre régulièrement sur les 24 heures à la demande. $\text{NOUMOYth} = \text{VEP}/12$.

- NOUMOY : régime le plus proche que l'on puisse obtenir à partir d'une configuration possible de la nourricière.

b) La première option envisagée : (test 1 changement)

Au plus un changement de régime nourricier

b.1) Les conditions d'accès à cette stratégie :

Pour pouvoir fixer un régime élévatoire nul pendant les périodes de pointe, sans changement de régime nourricier, il faut que l'apport d'eau de cette nourricière pendant les deux heures des périodes de pointe ne fasse pas déborder le réservoir d'effacement.

Ceci se traduit par l'équation :

$$2*NOUMOY < WEFFACMAX$$

D'autre part l'étude des courbes d'évolution des réservoirs de première élévation montre qu'il existe deux heures critiques (11 heures et 20 heures) pour lesquelles il faut s'assurer que la satisfaction de la demande en eau est possible. Ceci se traduit par les équations suivantes :

Pour la pointe de 11 heures :

$$DEM(7-11h) < \underset{\text{réserves}}{WELEV(7h)} + 2*NOUMOY + \underset{\text{apport maximal élévatoire}}{WEFFAC(7h)}$$

Pour la pointe de 20 heures :

$$DEM(7-20h) < \underset{\text{réserves}}{WELEV(7h)} + 11*NOUMOY + \underset{\text{apport maximal élévatoire}}{WEFFAC(7h)}$$

b.2) Le profil du régime nourricier proposé :

Dans ce cas on suppose que le changement est effectué à 7 heures du matin, et que la nourricière garde le régime NOUMOY jusqu'au lendemain matin.

c) La seconde option envisagée : (test 2 changements)
Deux changements de régime nourricier

c.1) Les conditions d'accès à cette stratégie :

Les conditions d'accès sont la satisfaction de la demande en eau à 11 heures (la même équation que précédemment). Mais cette fois, l'option NOUMOY constant ferait déborder le réservoir d'effacement pendant les heures d'arrêt du fonctionnement de l'élévatoire, ce qui se traduit par :

$$2*NOUMOY > WEFFACMAX.$$

Cette famille de stratégies sera aussi explorée au cas où aucune stratégie de la première famille n'est acceptable.

c.2) Le profil du régime nourricier proposé :

Dans cette option, les régimes successifs de la nourricière seraient calés en prenant comme hypothèse les heures de changement de régime fixées à 9 et 20 heures.

d) La troisième option envisagée : (test 4 changements)
Quatre changements de régime nourricier

d.1) Les conditions d'accès à cette stratégie :

Cette option sera envisagée si:

- $2 * \text{NOUMOY} > \text{WEFFACMAX}$
- la demande ne peut être satisfaite à 11 ou 20 heures,
- les options 1 et 2 ont été étudiées et aucune stratégie acceptable n'a pu être définie.

d.2) Le profil du régime nourricier proposé :

Les changements de régime de la nourricière suivraient dans cette hypothèse les changements de régime de l'élévatoire, c'est à dire seraient programmés à 9, 11, 18 et 20 heures.

G) Description des modules du programme PILOTE

1) Fonctionnalités du programme :

Ce programme commence par prendre en compte les indications données par les bases de connaissances, sur les seuils, objectifs, et classifications des critères. Il définit ensuite le régime nourricier en fonction de l'hypothèse donnée par la base. Enfin il génère et simule les stratégies partielles pour la première plage horaire.

Le programme génère, simule et sélectionne des stratégies partielles pour chaque famille dans chaque plage horaire.

Il se termine, en connectant, pour chaque famille, la stratégie sélectionnée par plage horaire.

Le programme est composé de différentes procédures incluses dans le programme principal. Chaque procédure remplit une des fonctions définies précédemment.

2) Le module de génération des stratégies (GENERATI)

a) Fonctionnalités

- Génération des stratégies d'une plage horaire en fonction de tests sur les débits productibles par chacune des configurations.

b) Entrées :

- Pour les autres plages horaires que la première, les fichiers BILAN..., créés par le programme de simulation sur l'étude de la plage horaire précédente.

c) Organisation du module

c.1) Les objets du module

Chaque objet stratégie est généré en cours de session et possède comme propriétés des numéros de plage horaire, de famille, de configuration, un indice de qualification pour chacun des 7 critères retenus, et sa potentialité est initialisé à "true" Celle-ci sera mise à "false", si elle franchit un seuil au cours de la phase de simulation.

Un exemple d'objet stratégie avec ses propriétés est donné ci-dessous :

STRATEGIE : no_plage : 1.

no_famille : 3
no_configuration : 9,
securité-EAU
securite-EDF
pression., etc....
potentialité : true.

c.2) La démarche du calcul

La règle de calcul des débits productibles donne une première approximation du volume productible par les configurations sur la durée de la plage horaire.

L'initialisation du choix des configurations est faite en calculant le volume à produire sur la plage horaire, et l'encadrement de ce volume compte-tenu, d'une part d'une fourchette fixée dans le réservoir de Montreuil, et d'autre part des limites du réservoir d'effacement.

Les configurations sont alors passées en revue les unes après les autres pour voir si elles satisfont à ce simple critère.

Une fois ces tests effectués, la procédure passe au programme de simulation les paramètres suivants:

- les objets stratégies générées avec leurs propriétés,
- les limites de la plage horaire,
- les valeurs initiales des niveaux des réservoirs,
- le nombre de stratégies générées.

3) Le module de simulation (SIMULA)

a) Fonctionnalités

Ce programme détermine, heure par heure, et pendant la durée de la plage horaire les niveaux prévisionnels des réservoirs d'effacement, du réservoir de Montreuil et la puissance électrique potentiellement appelée, pour chacune des configurations générées par le module de génération de stratégies.

Ce programme examine ensuite si ces niveaux et la puissance électrique appelée sont dans les limites des seuils définis dans la base de règles. Dans le cas de dépassement d'un seuil, la stratégie est éliminée. L'ensemble des niveaux et puissances sur les pas de temps horaires sont enregistrés dans un fichier appelé "BILAN-N°-stratégie". A toutes les stratégies potentielles simulées sur une plage horaire correspond donc un

fichier retraçant l'évolution prévisionnelle de la situation engendrée par cette stratégie.

b) Entrées

- Le fichier DEMANDE indiquant les prévisions de consommations,
- Le fichier DEBITNOU, créé par la base de règles REGINOUE, indiquant les 24 valeurs prévisionnelles du régime nourricier.
- Le fichier SURPRESS, permettant à partir de la connaissance de la configuration, de retrouver les corrections de pression à effectuer, dues au couplage des pompes.
- le fichier CONSELEC indiquant les consommations électriques sur l'usine autres que celles de la nourricière et élévatoire.
- les limites de la plage horaire,
- les seuils (seuil_R1, _R3, seuil_montreuil, seuil_électrique),
- les niveaux initiaux des réservoirs r1, r3, montreuil,
- les propriétés liées aux objets stratégies (numéros de la plage horaire, famille et configuration).

c) Sortie

- Un fichier BILAN.. par stratégie étudiée, gardant l'évolution prévisionnelle des niveaux et de la puissance électrique.

4) Le module de sélection

a) Fonctionnalités

La sélection de 2ème niveau intervient dans le cas où, dans une famille, plusieurs solutions ont été générées sur la plage horaire étudiée, puis simulées et non éliminées par les seuils contenus dans le module de simulation. Cette sélection a pour fonction de ne retenir dans la famille que la stratégie dont les caractéristiques sont les plus proches des objectifs visés dans cette situation d'exploitation. Cette sélection limite de fait le nombre de racines à partir desquelles vont être générées des stratégies pendant les plages horaires 3 et 4.

En cas d'impasse, c'est à dire s'il ne reste plus de solutions, un backtracking remettant en cause le choix du régime nourricier est envisageable à ce niveau, mais il n'a pas été organisé sur la maquette.

b) Entrées

- Les fichiers BILAN.. générés par le programme de simulation.

c) Le déroulement

Comme il s'agit de retenir une stratégie par famille, l'ensemble des stratégies d'une famille est passée en revue, au regard du ou des critères de sélection définis par la situation d'exploitation.

Le module est terminé quand toutes les stratégies ont été passées en revue.

5) Le regroupement

Ce module construit un fichier complet en connectant les fichiers BILAN, sélectionnés dans chaque plage horaire. Il affecte ensuite une qualification pour toutes les stratégies complètes, sur chaque critère.

6) Le classement

Ce module classe les stratégies complètes en fonction de l'algorithme de décision lexicographique.

H) Les fichiers de PILOTE

Dans la maquette, il y a trois types de fichiers :

- les fichiers contenant des données sur le fonctionnement de l'usine,
- les fichiers où sont stockées les résultats,
- les fichiers intermédiaires.

Lorsque le nombre de valeurs à passer était important, une transmission par un fichier intermédiaire a été préférée à un échange dynamique.

L'ensemble des fichiers de PILOTE est décrit dans le tableau ci-dessous.

nom	programme générateur	programme utilisateur	nb de champs	champs	type
PROFILS	*****	previsio	8	heure lundi mardi dimanche	ent " " "
(id pour été et printemps-automne.)					
PRE+date	previsio	*****	2	heure conso	ent "
DEMANDE	previsio	reginou	2	heure demande	ent "
SEUIL-OBJ	Contexte	PILOTE	8	seuils objectifs	réel "
DEBITNOU	reginou	simula	2	heure regime	ent "
SURPRESS	*****	simula	5	numéro nbE1 nbE11 nbE7 surpre	int " " " "
CONSELEC	*****	simula	4	heure pusine pozone pedf	int " " "
BILAN1	simula	regroupement	8	heure config pref niv1 hmtae vae niv3 hmtne vae mont Energae Energne Pdispo Coût	ent ent réel " " ent réel " " ent réel ent " " réel

Ch VI : L'EVALUATION DU SYSTEME

* * * * *

I) LE ROLE ET L'ORGANISATION DE CETTE PHASE

A/ Une évaluation et non une validation :

Le projet PILOTE a été défini comme un projet de "recherche et développement" dont la thèse constitue la partie recherche. Ceci signifie que l'architecture du logiciel ne résulte pas du travail d'un bureau d'études, appliquant des techniques confirmées (comme la recherche opérationnelle), mais d'un travail d'exploration de voies nouvelles.

La recherche méthodologique a été située, dès la définition de la thèse, dans le champ de l'intelligence artificielle. En particulier, il s'est agi d'introduire des éléments basés sur l'expérience et la connaissance de l'usine, pour guider et focaliser un algorithme de recherche.

Le logiciel précédemment implémenté a donc pour principale fonction d'illustrer une approche méthodologique dans la gestion concrète d'une usine, et c'est avec ce regard qu'il faut comprendre la phase d'évaluation.

C'est donc une maquette qui a été évaluée et la validation du système PILOTE ne pourra être envisagée que beaucoup plus tard, lorsque le système sera devenu opérationnel.

B/ Organisation de la phase d'évaluation :

PILOTE résulte d'un assemblage de plusieurs modules et raisonne en décomposant la journée en plages horaires. Aussi son évaluation a suivi ce double découpage.

Une première partie a consisté à tester séparément les différents modules. Ensuite les enchainements ont été évalués sur une seule plage horaire, puis sur la journée complète.

1) Une première évaluation module par module :

Les modules constituant l'architecture de PILOTE ont différentes fonctions et ne peuvent pas être tous évalués de la même façon. Le principe général est de se centrer sur un module, en maintenant constants les paramètres qui n'interviennent pas dans ce module.

Les modules à évaluer sont les suivants :

- la prévision de consommation,
- le modèle de simulation,
- les heuristiques de génération de stratégies,
- les heuristiques de sélection et de classement.

Une fois que les modules auront été testés séparément, il faudra étudier globalement les décisions prises par PILOTE d'abord sur une plage horaire, par enchaînement des différents modules, puis lors de sessions complètes.

2) De la plage horaire à la journée complète :

Le planning suivi pour tester les enchainements de raisonnements a été le suivant :

- les plages horaires de l'après-midi (11h-18h), du lundi 15 au vendredi 19 février.
- les plages horaires de nuit (20h,7h), du lundi 22 au vendredi 25 février.
- les enchainements après-midi-nuit (11h-7h), sur les périodes précédentes.
- les journées complètes, sur les périodes précédentes et du lundi 28 mars au vendredi 1 avril.

II) L'EVALUATION DES MODELES

A) Evaluation du modèle de prévision de consommation

1) Les deux aspects de l'évaluation

Le premier aspect est la précision, définie comme l'écart entre la prévision et la réalisation. Cet aspect a été étudié pour les prévisions journalières, pour les prévisions horaires et pour des prévisions correspondant aux périodes de planification. La précision a été étudiée, pour l'hiver, le printemps et l'été.

Le second aspect concerne l'influence des erreurs de prévision sur les décisions de choix de stratégies. L'idée était de rechercher la précision avec laquelle il est nécessaire de construire les modèles, pour ne pas altérer le processus décisionnel. Compte-tenu de l'état de développement du projet, cet aspect n'a été traité que pour 20 demi-journées de la période d'hiver, mais en prenant en compte les prévisions de 2 modèles.

2) Les modalités de l'étude

Une prévision de consommation horaire sur les 24 heures à venir a été réalisée par le modèle PROFIL sur l'usine de Neuilly d'une part entre le 15 et le 28 février pour la saison d'hiver, d'autre part entre le 11 avril et le 31 mai pour le printemps et enfin entre le 1 juin et le 31 juillet pour l'été.

Cette prévision est réalisée par un programme écrit en Turbo-Pascal "PREVISIO". Ce programme lit les informations sur les tirages horaires du jour précédent sur la disquette appelée "TRANSFERT PILOTE", alimentée par SOPHIA, le logiciel qui gère la base de données de l'usine. Ce programme se réfère aux prévisions de profils de consommations quotidiens contenus dans le fichier "PROFILS".

Les trois premières lettres du mois, les deux premières du jour de la semaine, le jour du mois et la température maximale prévisionnelle du jour sont entrées au clavier.

Après avoir reçu ces informations, le programme crée un fichier de nom "PRE-mois-jour" contenant les prévisions de tirage horaire, et un autre fichier de nom "ANA-mois -jour" contenant d'une part les prévisions et réalisations de la veille et d'autre part les informations sur les évolutions de niveau des réservoirs et sur les débits aux principales intercommunications (Villetaneuve et Choisy). Ce second fichier a pour rôle de faciliter l'analyse ultérieure des erreurs de prévision. C'est l'ensemble de ces fichiers "ANA" qui a servi de base à l'analyse des prévisions.

Par exemple, le 12 avril, le programme PREVI a lu le fichier APR11(tirages réalisés la veille) sur la disquette SOPHIA, et créé les fichiers PREAPR12 (contenant les prévisions du jour) et ANAAPR11 (analysant les erreurs de prévisions de la veille) sur le disque dur de l'IBM "PILOTE".

Le fichier PREAPR12, sera dupliqué sous le nom de CONSOPRE, qui servira de base à la constitution du fichier DEMANDE, utilisé par les modules de PILOTE.

Tous les aspects concernant la précision de la prévision ont été étudiés sur trois périodes :

- du 15 février au 28 février, correspondant à une période d'hiver,
- du 11 avril et le 31 mai, appelée plus loin dans le texte printemps,
- du 1 juin au 31 juillet, appelée plus loin dans le texte été.

Pendant cette période, une prévision de consommation horaire a été réalisée et enregistrée. Sur les 113 prévisions effectuées, 7 ont données des valeurs totalement aberrantes, surtout au mois de juin (15, 17, 22, 23, 24 juin et 3 juillet), et ces valeurs ont été retirées de l'analyse. Aucune explication n'a pu être donnée à ces prévisions aberrantes.

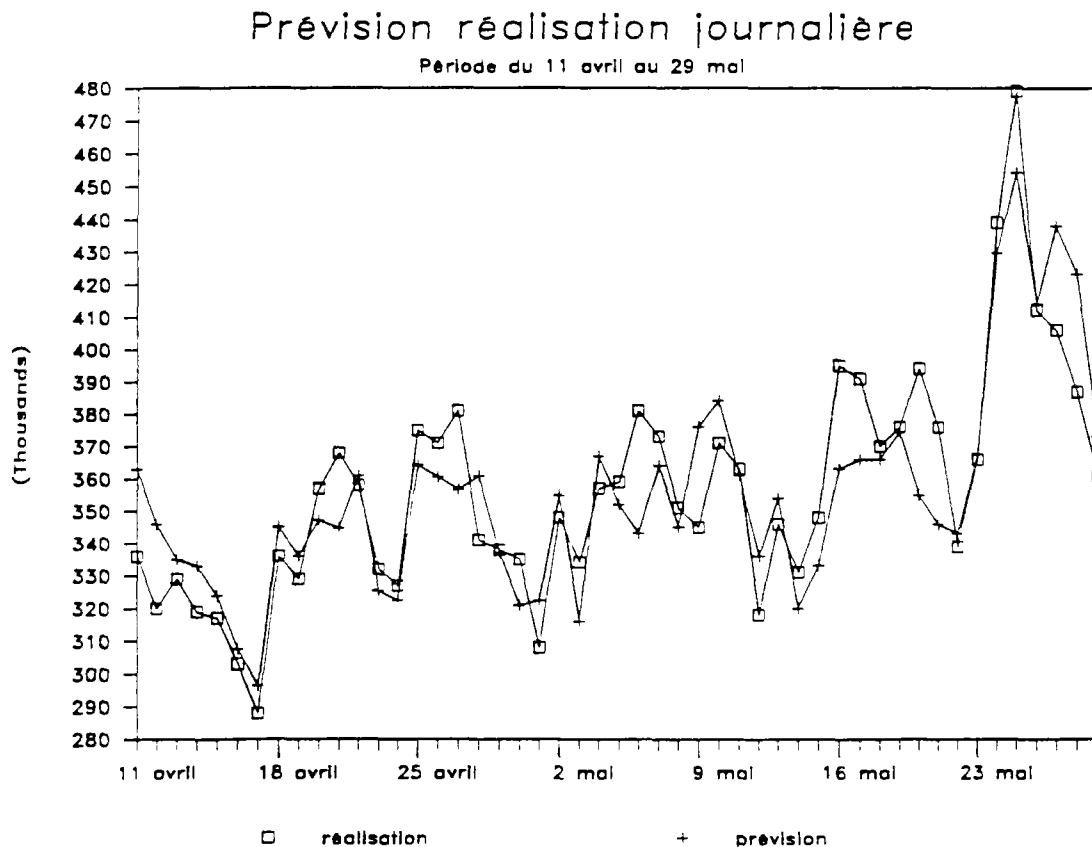
L'incidence des erreurs de prévisions sur la gestion a été étudiée pendant la phase d'évaluation de PILOTE du 15 février au 28 février.

3) Précision de la prévision

a) Analyse des erreurs de prévision sur les quantités journalières

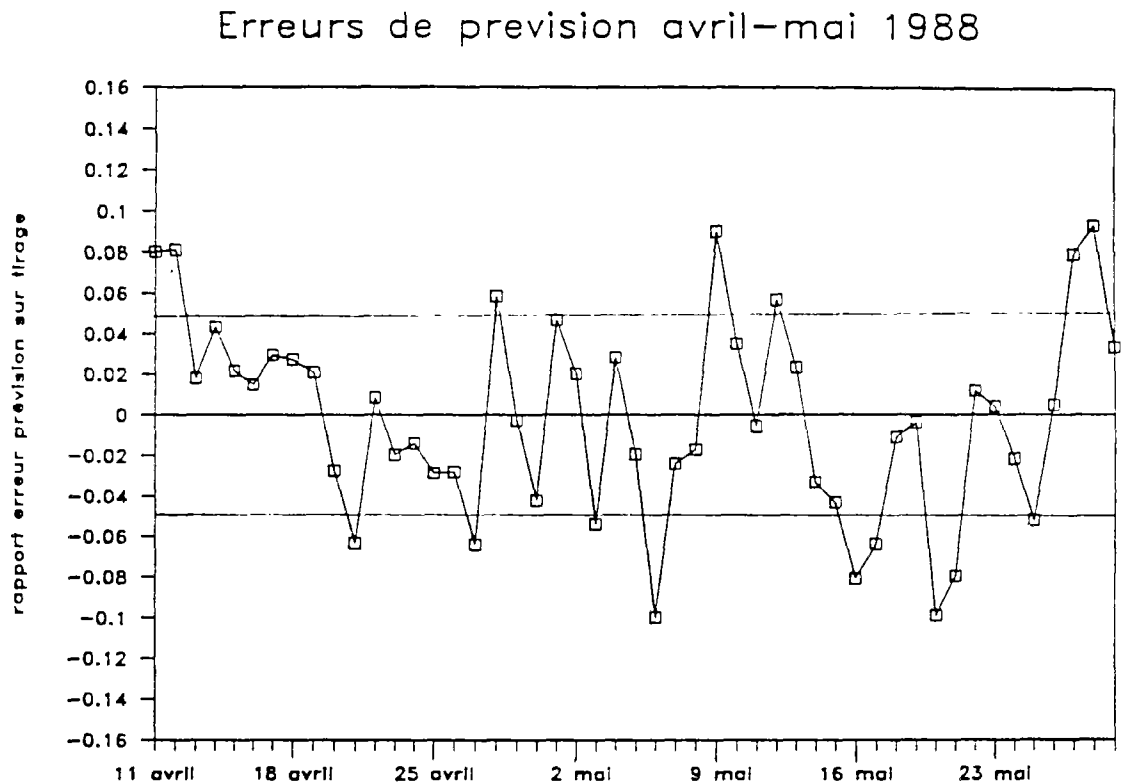
Ces résultats ont fait l'objet d'un document [LANNUZEL, PARENT 88], dont les principaux résultats sont donnés, en prenant exemple sur le printemps.

Le graphique suivant représente l'évolution comparée des prévisions et des réalisations pendant le printemps. Le tirage pendant cette période oscille entre 280000 m³ et 480000 m³.



Ce graphique indique que les erreurs de prévision sont comprises dans cette période dans la fourchette +/- 40000 m³. La proximité des deux courbes montre qu'il y a peu d'écarts entre la prévision de consommation journalière et la réalisation.

L'étude s'est prolongée par le calcul des rapports des erreurs de prévision sur le tirage. Le graphique ci-dessous visualise ces rapports pour le printemps.



1/ L'écart moyen du modèle journalier est $E_{alg} = -0,11 \%$, ce qui indique qu'il n'y a pas eu, sur cette période de tendance à la surestimation ou la sous-estimation.

2/ Les deux tiers des prévisions quotidiennes (35/51) sont dans une fourchette de précision de $+ 5 \%$. Deux prévisions sont erronées de plus de 10% (le 6 et le 20 Mai).

3/ La moyenne des valeurs absolues de ces rapports sur cette période correspond à un pourcentage d'erreur absolue $E_{abs} = 5,8\%$.

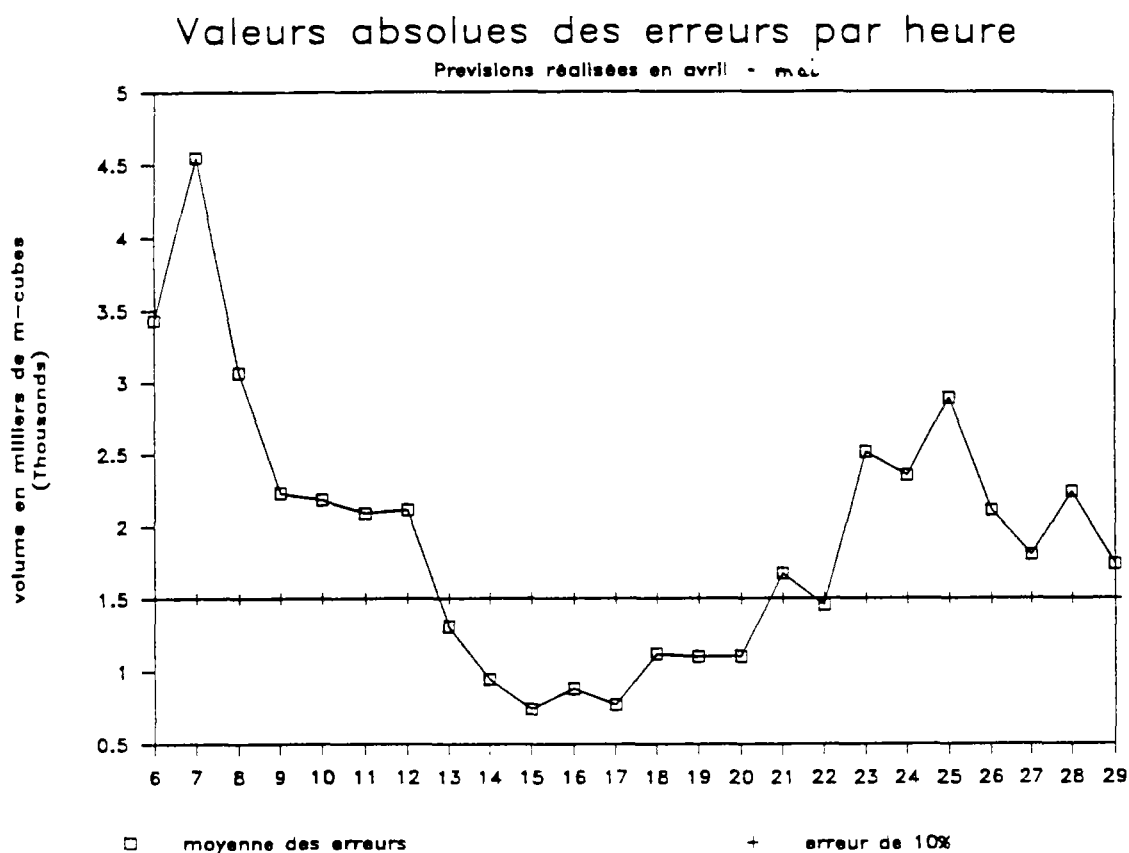
4/ On peut enfin remarquer qu'il y a des périodes continues de 5 à 9 jours pendant lesquelles le modèle, a surestimé ou sous-estimé, par exemple :

- surestimation du 11 au 19 avril, du 26 au 31 mai
- sous-estimation du 3 au 8 mai, ou du 14 au 21 mai

b) Analyse des erreurs de prévision sur les consommations horaires :

La moyenne des pourcentages d'erreurs absolues sur l'heure, calculée sur la période avril-mai, est de l'ordre de 16 %, et une analyse plus fine met en évidence l'inégale répartition des erreurs tout au long de la journée.

Le graphe ci-dessous représente les moyennes des erreurs absolues des prévisions horaires réalisées pendant le printemps 88.



Il apparaît nettement que les erreurs de prévision les plus importantes sont faites les 3 premières heures de la matinée :

6 h - 7 h : erreur absolue moyenne de 3 500 m³, soit de l'ordre de 23 %

7 h - 8 h : erreur absolue moyenne de 4 500 m³, soit de l'ordre de 23 %

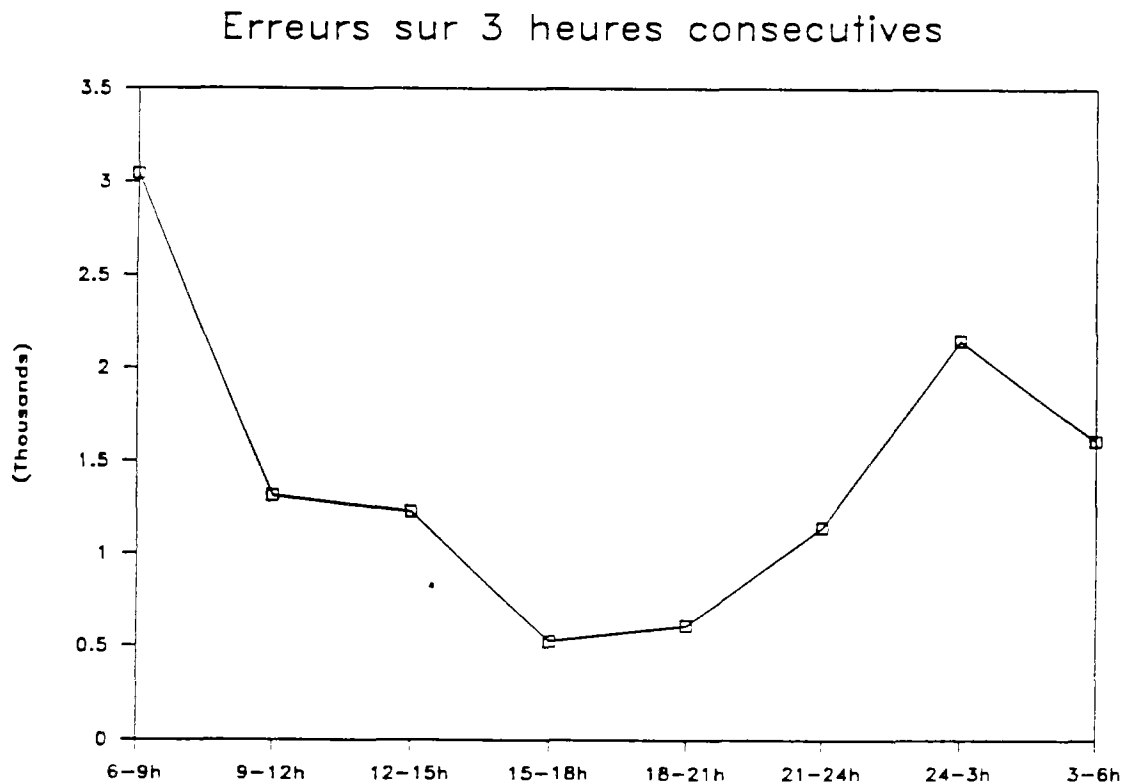
8 h - 9 h : erreur absolue moyenne de 3 000 m³, soit de l'ordre de 15 %

En fin de matinée et pendant les nuits, les moyennes oscillent entre 2 000 et 2 500 m³ avec une pointe à une heure du matin, soit de l'ordre de 20 %. Les prévisions les meilleures sont obtenues l'après-midi, pour la période comprise entre 13 et 20 heures.

c) Erreurs sur des séries consécutives de 3 et 8 heures :

La moyenne des pourcentages des erreurs absolues sur 3 heures consécutives est de 12 %, et de l'ordre de 9 % pour la durée du poste, ce qui indique une compensation des erreurs faites sur le pas de temps horaire.

Le graphe ci-après indique que les erreurs les plus importantes sont localisées en début de matinée, et dans une moindre mesure en début de nuit, ce qui va dans le même sens que les erreurs sur le pas de temps horaire.



Le tableau suivant illustre la décroissance des erreurs absolues avec l'augmentation de la durée de la prévision, ce qui indique bien une compensation des erreurs d'une heure sur l'autre quelque soit la saison. Le nombre N indique le nombre d'observations.

Durée	* N	* 1 heure	* 3 heures	* 8 heures	* 24 heures
fevrier	* 336	* 10,4 %	* 6,9 %	* 3,4 %	* 2,0 %
printemps	* 1200	* 16,9 %	* 12,4%	* 8,9 %	* 5,4 %
été	* 1200	* 14,2 %	* 11,1 %	* 9,2 %	* 6,4 %

c) Bilan de l'aspect précision

Des prévisions ont été réalisées en continu sur le site depuis le 11 avril correspondant à 2 mois de la période de printemps et 2 mois de la période d'été.

Un certain nombre de problèmes se sont posés que l'on peut classer en 2 catégories :

- les difficultés indépendantes du modèle de prévision
- les difficultés inhérentes au modèle ou à son utilisation.

c.1) Les difficultés indépendantes du modèle :

Pendant les 4 mois d'utilisation, on a pu observer plusieurs cas d'erreurs dans la transcription des données entre SOPHIA et PILOTE. Les fichiers contenaient des données en totalité ou en partie erronnées. Ces erreurs relevaient apparemment de problèmes liés à la tête de lecture-écriture de l'IBM sur lequel est implanté SOPHIA.

Rappelons que pendant 7 journées de la fin du mois de juin, le calcul des prévisions a donné des résultats aberrants alors que les fichiers de données étaient corrects, et il n'a pas été possible de trouver une explication à ces aberrations qui ne se sont pas reproduites ensuite.

Notons enfin que les données en provenance de Sophia sont entachées des erreurs de saisie qui n'ont pas été corrigées.

c.2) Les difficultés inhérentes au modèle ou à son utilisation :

Deux cas rencontrés pendant la période testée illustre bien ce type de difficultés.

Le 10 juin, l'usine a fourni une quantité d'eau importante à la ville de PARIS, quantité évidemment non prévue par le modèle et qui n'a pas été retirée du calcul de la consommation. Aussi à l'"erreur" du 10 juin se serait répercutée les jours suivants, si la consommation n'avait pas été traitée avant d'être validée pour la suite.

Le jeudi 14 juillet n'a pas été considéré comme un jour férié mais comme un jour de semaine ordinaire, ceci a entraîné une première erreur de prévision pour ce jour là, erreur qui s'est propagée pendant quelques jours.

Notons enfin que la température maximale prévisionnelle du jour a le plus souvent été donnée au jugé, et non demandée aux services météorologiques de la région.

d) Quels remèdes :

Ces quelques cas mettent en évidence la nécessité d'un suivi quotidien des prévisions, dont les fonctions principales seraient :

- contrôle des données (toute erreur de saisie a des conséquences sur la prévision),
- regard critique sur le déroulement de la journée (au niveau des intercoms),
- prise en compte des particularités des jours (fériés, ponts, grèves...),

Ceci suppose la mise en place sur l'usine d'une structure légère chargée de gérer la prévision, en exerçant un contrôle sur les données des jours précédents servant de base aux prévisions.

4) Incidences des erreurs de prévisions sur les décisions

a/ Le principe de l'étude :

Les prévisions sont utilisées pour définir la meilleure gestion hydraulique de l'usine. Aussi l'idée directrice de ce chapitre est la comparaison des stratégies définies à partir de prévisions avec les stratégies qui auraient été choisies en toute connaissance des consommations. Cet ensemble de stratégies est bien sûr défini a posteriori.

Pour pouvoir tester la sensibilité des mécanismes de prise de décisions aux aléas de la prévision, la période de planification retenue a été la plage horaire, qui contrairement à la journée complète, permettait de quantifier l'impact des erreurs de prévision sur le mécanisme de prise de décision.

En effet, sur la plage horaire, il est encore possible de voir quelle modification dans la qualification d'un critère pour les stratégies, a conduit à une décision différente.

b/ Niveaux d'intervention de la prévision dans la prise de décision :

Dans le processus de décision de PILOTE, la prévision de consommation intervient dans deux modules différents :

- dans le premier filtrage quantitatif fait avec un modèle sommaire,
- dans le module de simulation, car l'ensemble des évolutions prévisionnelles est calculé à partir d'une modélisation du tirage.

Les erreurs de prévision peuvent donc avoir trois types de conséquences :

- éliminer d'emblée des stratégies qu'il aurait été intéressant d'examiner, ou décider de simuler des stratégies a posteriori inutiles,
- éliminer des stratégies pendant la phase de simulation pour cause de franchissement d'un seuil,
- modifier la qualification des stratégies sur un ou plusieurs critères et donc choisir une autre stratégie en suivant la procédure d'examen des critères définie.

c/ Illustration de la comparaison sur une journée particulière :

Les tableaux C.3.1. et C.3.2. contiennent respectivement les stratégies qui ont été générées, d'une part a priori avec les prévisions du modèle profil, et d'autre part a posteriori avec les données historiques de consommation.

Stratégies générées à priori le 23 février avec les prévisions du modèle PROFIL.

Indice	20	21	22	23	24	1	2	3	4	5
1	111	--	--	--	--	--	--	--	--	--
2	112	--	--	--	--	--	--	--	--	--
3	120	--	--	--	--	--	--	--	--	--
4	121	--	--	--	--	--	--	--	--	--
5	111	--	--	--	--	112	--	--	--	--
6	111	--	--	--	--	120	--	--	--	--
7	111	--	--	--	--	121	--	--	--	--
8	112	--	--	--	--	120	--	--	--	--
9	112	--	--	--	--	121	--	--	--	--
10	120	--	--	--	--	112	--	--	--	--
11	120	--	--	--	--	121	--	--	--	--
12	121	--	--	--	--	112	--	--	--	--
13	121	--	--	--	--	120	--	--	--	--

Stratégies générées à posteriori pour le 23 février en utilisant les données historiques.

Indice	20	21	22	23	24	1	2	3	4	5
1	111	--	--	--	--	--	--	--	--	--
2	112	--	--	--	--	--	--	--	--	--
3	120	--	--	--	--	--	--	--	--	--
4	121	--	--	--	--	--	--	--	--	--
5	111	--	--	--	--	112	--	--	--	--
6	111	--	--	--	--	120	--	--	--	--
7	111	--	--	--	--	121	--	--	--	--
8	112	--	--	--	--	120	--	--	--	--
9	112	--	--	--	--	121	--	--	--	--
10	120	--	--	--	--	112	--	--	--	--
11	120	--	--	--	--	121	--	--	--	--
12	121	--	--	--	--	112	--	--	--	--
13	121	--	--	--	--	120	--	--	--	--

La comparaison des deux tableaux indique que des stratégies identiques ont été générées par les deux séries de données.

La seconde série de tableaux contient les qualifications données aux différents critères après la simulation des stratégies.

Tableau des qualifications aux différents critères pour les stratégies simulées avec les prévisions du modèle PROFIL et les données historiques (repérées par '):

Stratégie	secumon	secuedf	pression	dist-M	dist-E	nb-chgt	cout
1	moy	bon	bon	mauv	mauv	0	19800
1'	--	--	--	--	--	--	--
2	moy	bon	bon	moy	mauv	0	21400
2'	--	--	--	mauv	--	--	21450
3	moy	bon	bon	bon	mauv	0	23200
3'	--	--	--	mauv	--	--	23300
4	moy	bon	moy	bon	bon	1	24700
4'	--	--	--	--	--	--	24800
5	moy	bon	bon	mauv	mauv	1	20600
5'	--	--	--	--	--	--	20600
6	moy	bon	bon	moy	mauv	1	21500
6'	--	--	--	--	--	--	21600
7	moy	bon	bon	bon	mauv	1	22400
7'	--	--	--	--	--	--	22400
8	moy	bon	bon	bon	mauv	1	22300
8'	--	--	--	moy	--	--	22400
9	moy	bon	moy	bon	mauv	1	23100
9'	--	--	--	--	--	--	23200
10	moy	bon	bon	bon	mauv	1	22400
10'	--	--	--	moy	--	--	22400
11	moy	bon	moy	bon	moy	1	24000
11'	--	--	--	--	--	--	24000
12	moy	bon	bon	bon	mauv	1	23100
12'	--	--	--	moy	--	--	23100
13	moy	bon	bon	bon	moy	1	24000
13'	--	--	--	--	--	--	24000

Un examen attentif met en évidence des différences de qualification des critères "dist-Mont" et "dist-effac", correspondant aux écarts aux objectifs fixés pour le réservoir de Montreuil et le réservoir d'effacement. Par exemple la stratégie n° 10 a comme qualification pour le critère "dist-Mont", bon avec les données du modèle, et moyenne avec les données historiques.

Cependant, et compte-tenu du processus de sélection des stratégies décrit dans le rapport "Ajustement et évaluation du logiciel

PILOTE" (mai 88) sur ce cas, c'est la stratégie n° 3 qui aurait été sélectionnée pour les deux séries de données.

4) Bilan sur les 10 périodes étudiées :

Pour cette étude, les prévisions d'un autre modèle ont été intégrés. Ce modèle, de type ARIMA, est décrit dans le rapport [LANNUZEL, PARENT 88]. Vingt périodes (10 après-midis et 10 nuits) ont été étudiées avec les prévisions du modèle PROFIL et du modèle ARIMA, ce qui représente un total de 40 cas.

La première étude a consisté à déterminer à partir de quel pourcentage d'écart "prévision-réalisation", les décisions étaient affectées.

Ensuite, un examen plus fin des causes des différences de choix, a été mené pour rechercher les critères les plus sensibles.

Le tableau suivant analyse les différences de décision, pour chaque classe de pourcentage d'écart, en indiquant, sur la première ligne, le nombre de prévisions de la classe, et sur la seconde, le nombre de stratégies différentes choisies.

Classe	- 2 %	2 à 3 %	3 à 4 %	4 à 5 %	5 à 6 %	6 à 7 %	+ 7 %
N pré	2	4	8	10	9	5	2
Ch. dif.	0	0	1	2	8	5	2

Il apparaît clairement que les conséquences des erreurs des modèles sont sensiblement différentes suivant que la différence "prévision-réalisation" sur la période testée est supérieure ou inférieure à 5 %.

A moins de 5 % d'écart, ce sont alors à quelques exceptions près les mêmes stratégies qui sont générées et la phase d'élimination affecte les mêmes stratégies.

Des différences sont observées au niveau de la qualification du critère "distance à l'objectif", mais les modifications ayant lieu dans le même sens pour l'ensemble des stratégies, le choix final n'est affecté que dans quelques cas. Et pour ces cas là, les stratégies sélectionnées en fonction des prévisions du modèle ont des caractéristiques voisines de celles qui auraient été choisies à postériori.

Dans le cas où les erreurs sont supérieures à 5 %, on observe un décalage dès la phase de génération des stratégies. Une surestimation du modèle conduit à

une génération de stratégies de pompage plus importantes, et inversement pour une sousestimation.

D'autre part, des stratégies communément générées par le modèle et à posteriori peuvent être éliminées dans un cas et pas dans l'autre.

Enfin, la procédure de sélection est toujours affectée et les stratégies retenues par le modèle, sont en conséquence, pratiquement toujours différentes. Il apparaît, par un examen plus fin, que le critère le plus sensible est la "distance à l'objectif de Montreuil".

Une prévision avec une précision inférieure à 5 % n'altérant pas sensiblement la procédure de définition de stratégies de pompage, c'est donc cette précision qu'il faut viser dans l'élaboration des modèles de prévision.

Remarquons enfin que, compte-tenu de la structure des erreurs de prévisions, il apparaît que les décisions sont peu affectées pour les périodes de planification correspondant à l'après-midi, dans un peu plus de la moitié des cas pour la nuit, et pratiquement toujours pour la matinée.

B) Le modèle de simulation

1) Le principe de l'évaluation de ce modèle

Le modèle a été construit à partir d'une description fonctionnelle des aspects hydrauliques et énergétiques de l'usine de Neuilly.

Un modèle de simulation est par nature déterministe. Ce qui veut dire que la connaissance de toutes les données et de l'ensemble des phénomènes physiques, devrait conduire à une description parfaite de la réalité et donc à des erreurs négligeables. Mais tout modèle est obligatoirement basé sur des simplifications et est donc par nature inexact. Les approximations faites dans le modèle ont été décrites dans le chapitre 1.

L'évaluation du modèle de simulation a consisté à recenser les erreurs ou dérives, à déterminer leur ordre de grandeur, à analyser leurs conséquences sur le raisonnement et enfin à étudier les possibilités d'améliorer la précision du modèle.

Pour tester la fidélité du modèle de simulation par rapport à la réalité de l'usine, les évolutions observées de niveaux et de consommations électriques ont été comparées avec celles calculées par le modèle à partir des valeurs historiques correspondantes des débits, tirages. L'exactitude du modèle a été évaluée par un suivi en temps réel de ses dérives, pas de temps par pas de temps.

2/ Résultats de l'évaluation du modèle de simulation

a/ La pression du refoulement

L'analyse des résultats montre que l'écart entre la pression calculée par le modèle et la pression relevée heure par heure par les chefs de poste a rarement dépassé 1 m. sur une pression moyenne de 95 m.

L'influence de cet écart sur le débit refoulé peut être évaluée à partir du calcul d'incertitude.

La relation débit-hauteur manométrique est une équation de la forme :

$$Q = a h^2 + b.h + c$$

$$dQ = 2.a.h dh + b dh$$

L'incertitude ΔQ est liée à l'incertitude ΔH par :

$$\Delta Q = (2.a.h.dh + b) . \Delta h$$

Pour la hauteur manométrique moyenne $h = 90$ m., on obtient pour les trois familles de pompes; pour une incertitude $\Delta h = 1$ m.

Pompes de type E₁ - E₆ :

$$\Delta Q = 63. \Delta h \text{ pour } q = 5600 \text{ m}^3, \text{ soit } \Delta Q/Q = 1,2 \%$$

Pompes de type E₇ - E₈

$$\Delta Q = 280. \Delta h \text{ pour } q = 10900 \text{ m}^3, \text{ soit } \Delta Q/Q = 2,5 \%$$

Pompes de type e₁₁ - e₁₂

$$\Delta Q = 25. \Delta h \text{ pour } q = 1800 \text{ m}^3, \text{ soit } \Delta Q/Q = 1,4 \%$$

Un écart inférieur à 1m entre les pressions calculées et relevées induira donc une erreur relative sur le débit, maximale de 2,5 %.

On peut conclure que la modélisation sommaire de la pression de refoulement est compatible avec la précision avec laquelle le logiciel PILOTE peut travailler.

b/ Le calcul des volumes refoulés

Comme cela est décrit dans le modèle de simulation, le calcul des volumes refoulés par l'élévatoire est effectué à partir des équations utilisées dans SOPHIA et dans l'UCO.

Les volumes refoulés étant inconnus sur l'usine, l'adéquation de la modélisation à la réalité de l'usine ne peut être évaluée que par la comparaison des évolutions de réservoirs.

b.1/ Pour les après-midis d'hiver :

Pour les quatre premiers après-midis, les écarts entre les résultats des calculs entre les résultats des calculs du modèle et les relevés horaires de niveau restent faibles.

La dérive que l'on observe pour les quatre dernières heures de la dernière journée correspond à la mise en route à 14 h des trois surpresseurs de Noisy II, qui ont entraîné une baisse de pression de 1,5 m, donc une augmentation du débit refoulé qui n'a pas été prise en compte par le modèle.

Cela se traduit concrètement par une surestimation du volume dans le réservoir d'effacement, compensée par une sous-estimation du niveau de remplissage de Montreuil.

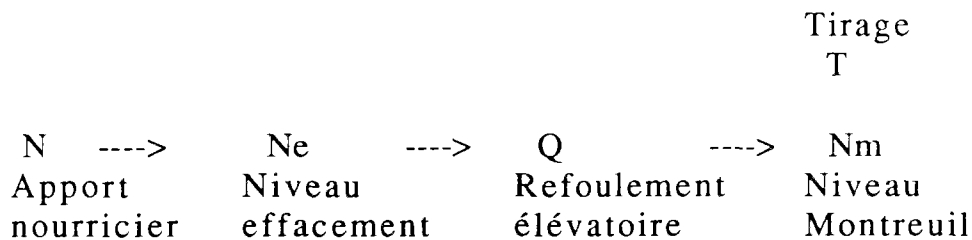
b.2/ Pour les nuits d'hiver

Pour les quatre nuits étudiées, on observe une dérive identique et symétrique des évolutions des réservoirs de Montreuil et d'effacement. Tout se passe comme si le modèle surestimait les débits refoulés, et donc, transférait l'eau du réservoir d'effacement à celui de Montreuil.

L'écart entre les niveaux calculés et relevés en fin de nuit correspond à un volume de l'ordre de 10.000 m³, soit un peu moins de 10 % du volume total pendant la nuit, et le calcul indique que l'excès à Montreuil est égal au volume manquant dans l'effacement à quelques centaines de m³ près.

b.3/ Recherche d'explication

En schématisant, l'évolution du niveau de l'effacement est lié à l'apport nourricier et au prélèvement élévatoire, et celle du niveau de Montreuil à l'apport élévatoire diminué du tirage.



Le problème se caractérise alors par un jeu de 3 inconnues Q, N et T, reliés par un jeu de 2 équations donnant les niveaux à Montreuil et dans l'effacement (Nm et Ne).

Plusieurs hypothèses sont envisageables pour expliquer la dérive.

Hypothèse 1 : la relation débit-hmt est exacte.

- La dérive dans le réservoir d'effacement ne serait explicable que par une sous-estimation de l'apport nourricier.

Compte tenu du régime moyen de la nourricière pendant les nuits, l'apport nourricier devrait être augmenté de 5 % pour combler le déficit ;

- La dérive à Montreuil ne serait explicable que par une sous-estimation du tirage, qui devrait être égale à la sous-estimation du tirage nourricier.

Hypothèse 2 : la relation débit-hmt est inexacte :

Une surestimation du débit refoulé expliquerait à la fois le déficit de l'effacement et l'excès à Montreuil, sans modifier l'apport nourricier et le tirage.

Cette surestimation semble d'autant plus sensible que les hauteurs manométriques augmentent, et que la proportion d'eau élevée par l'ancienne élevatoire est plus importante. Cela donne le modèle.

c) Le bilan

L'évaluation indique clairement les limites du modèle, et quelques indices pour l'améliorer.

III L'EVALUATION DES HEURISTIQUES DES BASES DE CONNAISSANCES

A) Le sens et la méthodologie de l'évaluation

Pilote contient plusieurs heuristiques, par exemple dans le module CONTEXTE, pour définir les critères de choix des stratégies que l'on va continuer à prospecter.

Les bases de connaissances ont été ajustées à partir des réponses aux questionnaires distribués aux chefs de poste. PILOTE doit donc assumer que ses heuristiques représentent les approches qui pourraient être suivies par les chefs de poste. Il est donc nécessaire de confronter les décisions de PILOTE aux experts pour vérifier que l'expertise représente bien leur jugement.

Dans d'autres travaux sur les systèmes experts [BONNET 86], la validation des heuristiques est généralement faite à partir de cas tests. Les experts voient comment le système réagit dans un grand nombre de situations et à partir de là évaluent la performance. La validation des bases de connaissances s'inspirera du même principe.

Pour évaluer chaque base de connaissance, un ensemble de cas-tests, particuliers à cette base, seront définis. Chaque cas-test sera caractérisé par un ensemble d'informations que PILOTE est censé connaître en amont du module en cours d'évaluation.

Ce cas test servira d'une part de base de faits et d'autre part sera présenté avec les mêmes informations aux chefs de poste. Ensuite les résultats donnés par les experts et le programme seront comparés.

B) La génération des stratégies

Les premières heuristiques, sur les éliminations de stratégies à priori, ont été reprises telles qu'elles ont été formulées par les chefs de poste, et validées par le directeur technique. Leur évaluation, vis à vis des décisions prises par les chefs de poste, est donc inutile.

C) Les heuristiques du premier niveau de filtrage

1) La question posée

La question posée dans l'évaluation de ce module est la suivante:

Est-ce que les stratégies générées par PILOTE correspondent aux stratégies que les chefs de poste envisagent, dans le contexte donné, et que "PILOTE" examinera plus en détail ?

2) Les modalités de l'étude

Pendant les 15 jours de la période d'évaluation d'hiver, il a été demandé aux chefs de poste, d'indiquer, en début de période de planification, quelles stratégies ils envisageaient. Ils avaient à leur disposition, les mêmes informations que "PILOTE", et en particulier, les prévisions du modèle. Il a ensuite été noté quelle stratégie avait été effectivement suivie dans la plage horaire.

Les études de cas sont constituées par les indications données en temps réel, par les chefs de poste, au moment de choisir une stratégie de pompage. Ces indications ont été obtenues à l'aide d'un questionnaire.

3/ Résultats de l'évaluation

a/ Pour les plages des après-midis d'hiver :

La comparaison des stratégies partielles générées par PILOTE, sur une semaine d'hiver, avec celles envisagées par les chefs de poste en début de plage horaire amène deux remarques.

Date	Stratégies Considérées par les chefs de poste		Nombre de Stratégies Générées par PILOTE (a)	Stratégies Considérées par les chefs et non Générées par PILOTE
22 Fev	120	121	12	
23 Fev.	120	121	13	
24 Fev.	112	120	10	
25 Fev.	112	120	9	

a. Toutes les stratégies sont générées avec les prévisions du modèle

Tout d'abord le nombre moyen (12) de stratégies générées par PILOTE est très supérieur à celui des stratégies envisagées qui est de 2. Cela vient du fait que les chefs de poste n'ont pas envisagé dans cette période test, de

changement de configuration dans la plage horaire, ce qui par contre a été le cas de PILOTE.

Ensuite, il est à noter que toutes les stratégies envisagées pour cette plage horaire sont bien incluses dans l'ensemble des stratégies générées par PILOTE.

b/ Pour les plages horaires de nuit :

La même comparaison met en évidence un plus grand nombre de stratégies envisagées, car dans ce cas, les chefs de poste prennent en compte les changements de configuration indiqués dans les heuristiques de génération.

Date	Stratégies Considérées par les chefs de poste	Nombre de Stratégies Générées par PILOTE (a)	Stratégies Considérées par les chefs et non Générées par PILOTE
22 Fev. 1988	120b 121 120/122 *120/121c	14	120/122
23 Fev. 1988	120 121 120/122 *120/121	13	120/122
24 Fev. 1988	112/121 *120/121	12	
25 Fev. 1988	120 120/121 *112/121d	13	120/121

a. Toutes les stratégies sont générées avec les prévisions du modèles

b. Quand une simple configuration est mentionnée, cela signifie qu'il n'y a pas de changement pendant la période.

c. Quand deux configurations sont mentionnées, cela signifie qu'il y a un changement pendant la période, dans ce cas à 1 h du matin.

d. La configuration 112/121 a été générée par PILOTE, mais éliminée pendant la simulation hydraulique.

Sur l'ensemble de la semaine testée, il apparaît aussi que pour 3 nuits, une stratégie partielle envisagée n'a pas été générée par PILOTE. Ces stratégies partielles manquantes ont comme point commun d'avoir été rejetées après le calcul sommaire de dégrossissage, pour dépassement de la limite supérieure de la fourchette.

Cela est un effet induit de la surestimation des volumes pompés par l'élévatoire.

L'ensemble des comparaisons indique qu'un ajustement d'expertise est à prévoir pour ces heuristiques de génération, particulièrement pour resserrer le filtrage, c'est à dire pour diminuer le nombre de stratégies générées.

D) Les heuristiques de sélection des configurations à prospecter

1) La question posée

Pour ce module, il s'agissait de répondre à :

Est-ce que la stratégie préférée par PILOTE dans l'ensemble des stratégies possibles dans une plage horaire est bien celle qu'aurait choisie les chefs de poste dans ce cas ?

2) Les modalités de l'étude

Les études de cas ont été basées sur la présentation d'un ensemble de stratégies partielles candidates. Les chefs de poste devaient indiquer celle qui leur paraissait la meilleure (et donc, par différence, celles que l'on élimine à la fin de cette plage horaire).

Les caractéristiques essentielles d'une stratégie partielle ont été présentées, en grande partie sous forme graphique et synthétique (une page). Un jeu de stratégies à sélectionner était composé de cinq ou six unités pour permettre un tri rapide de la part des chefs de poste.

L'observation des procédures de tri et l'analyse des différences entre les choix a servi à renforcer l'expertise.

3/ Résultats de l'évaluation des jeux d'heuristiques :

Les cinq études de cas qui ont servi à l'évaluation de ces heuristiques étaient composées d'un ensemble des 5 ou 6 stratégies partielles correspondant à des plages d'après-midis d'hiver.

Un examen attentif des résultats fait apparaître que les stratégies retenues ne sont pas identiques pour l'ensemble des chefs de poste, même s'ils suivent globalement une démarche voisine. De petites différences d'appréciation d'un critère conduisent parfois à un choix différent de stratégie de pompage. Aussi, est-il apparu plus judicieux de ne pas se focaliser trop sur les résultats, mais plutôt sur la démarche suivie.

L'observation de leurs processus de choix de stratégies a mis en évidence plusieurs points.

Les critères ont bien été pris en compte selon l'importance défini dans les réponses au questionnaire. Par contre, le choix de la stratégie préférée n'est pas basé sur un strict respect de cette hiérarchie et les critères ne sont pas utilisés linéairement.

Il y a en fait un regroupement des différents critères, par niveau, ce qui introduit une souplesse par rapport à la hiérarchie définie. Cet écart par rapport à ce qui est implémenté, a conduit à la remise en cause de la sélection lexicographique stricte. Un renforcement de l'expertise a orienté la démarche de sélection vers une technique d'agrégation de critères, qui est présentée plus loin.

IV) L'EVALUATION GLOBALE

A) L'évaluation des enchaînements

1) Un enchaînement technique satisfaisant, en tant que maquette

Tous les tests ont indiqué que le système enchaînait correctement les modules, sur une période de planification, et que la connection des différentes plages horaires se faisait correctement. Le seul problème rencontré est la fragilité du système. En tant que maquette, le système ne dispose pas de contrôle sur les données saisies, ni de procédures de récupération des erreurs de manipulation.

2) Des comparaisons difficiles avec les experts

Lorsque le système produit des stratégies complètes, l'objectif n'est plus de se calquer sur la démarche des chefs de poste, parce que l'ensemble du système de traitement de l'information, sur une journée, est trop complexe pour être appréhendé dans un raisonnement humain.

L'évaluation consiste plutôt à se référer aux objectifs définis lors de la conception du système. Ceci signifie tout d'abord vérifier que le système est bien modulaire, et qu'il permet de bien marquer la trace du raisonnement. Enfin, il convient de s'assurer que les informations sont utilisées à bon escient, pour réduire au mieux, à chaque étape, l'espace de représentation du problème.

B) L'évaluation des objectifs définis lors de la conception

1) La modularité du système

L'implémentation du système a bien respecté cet objectif. L'ensemble de la démarche a effectivement été décomposé en bases de connaissance ou

procédures, ayant chacune des fonctionnalités limitées, ce qui facilitera leur maintenance.

La séparation entre bases de connaissances et programmes, ne pose pas de problème d'enchaînement logique, mais un désagrément lors de l'exécution, parce qu'il faut charger (et décharger) en mémoire vive successivement, Turbo-Pascal, Lotus (Visualisation graphique), Nexpert, Turbo-Pascal et Lotus. La séquence est toutefois automatisée par un fichier Batch.

2) La mémorisation de la trace du raisonnement

Le système fonctionne concrètement sur un ensemble de filtres qui tamisent successivement les stratégies. Dans sa version actuelle, les stratégies générées apparaissent en cours de session, puis il est indiqué à l'écran, le moment de leur élimination. Même si cela n'est pas implémenté, il serait possible de transférer dans un fichier, le niveau de raisonnement auquel une stratégie a été éliminée, et la cause de son élimination.

De tels fichiers constitueraient une base de travail pour l'ajustement des objectifs, seuils et classes de qualification et modes de regroupement des critères.

3) L'efficacité de l'utilisation des informations

Les filtres mis en place se sont révélés suffisamment efficaces, pour que la réduction des stratégies envisageables à chaque niveau soit compatible avec une recherche d'informations plus précise au niveau suivant. En particulier, le nombre de stratégies partielles simulées sur une plage horaire, même s'il est beaucoup plus important que celui des stratégies envisagées par les chefs de poste, reste de l'ordre d'une centaine, ce qui est encore gérable par un micro-ordinateur.

Enfin, le raisonnement qualitatif, coûteux en temps de calcul (tests nombreux), ne concerne qu'un nombre limité de stratégies.

C) Une réflexion sur le renforcement des heuristiques de sélection

1) Une orientation de la réflexion vers l'agrégation de critères

Plusieurs auteurs [HART 88], [VOGEL 88], remarquent que les experts manipulent rarement simultanément plus de 2 ou 3 critères. Comme la plupart des problèmes de comparaison demandent de manipuler une dizaine de critères, l'expert opérationnalise l'information qu'ils contiennent en réduisant leur nombre en les agrégeant en un ou plusieurs niveaux

synthétiques. En effet, moins il y a de critères, plus la réalisation d'un choix est rapide et efficace.

Ce qui suit relève d'une réflexion méthodologique pour utiliser les critères. Cette réflexion est restée au stade initial, et n'a été ni implémentée, ni testée.

L'idée directrice de cette proposition, basée sur l'optimisation de l'utilisation de l'information, est de regrouper les critères en plusieurs paliers, de façon à définir un seul critère synthétique, qui soit à la fois représentatif, et redécomposable.

La volonté qu'il soit décomposable correspond à la nécessité de pouvoir traiter les cas où l'utilisation du seul critère synthétique ne suffit pas à extraire une et une seule stratégie de l'ensemble des stratégies candidates. Dans ces cas, il y a alors nécessité de rechercher des informations plus fines, en analysant les critères, au niveau synthétique immédiatement inférieur.

Comme à ce niveau, il y a plus de critères, la manipulation d'un grand nombre d'informations est bien sûr plus complexe. Mais l'on retrouve l'idée directrice qui veut que cette recherche n'est effectuée que pour le nombre limité de stratégies classées dans la meilleure catégorie du critère synthétique. Ce critère synthétique fonctionne en fait comme un nouveau filtre.

2) Présentation de la démarche envisagée

L'objectif de l'agrégation est de passer d'un ensemble de critères élémentaires à un seul critère global en les regroupant en différents niveaux.

Pour des questions de simplicité et de meilleure compréhension, l'exemple donné regroupe les critères 2 par 2.

Une agrégation de critères pose 2 types de questions. La première est de savoir comment associer les différents critères dans l'arbre du regroupement. La seconde est de définir la qualification du critère synthétique en fonction des qualifications respectives des 2 critères qu'il agrège.

3) La construction de l'arbre de regroupement

Lors de la phase de recueil de l'expertise, 7 critères élémentaires ont été définis, avec leurs classes respectives de qualification. Les propositions

de regroupement qui suivent, résultent d'entretiens avec les chefs de poste et les directeurs de l'usine.

a) Les regroupements de premier niveau

Ainsi, les critères "sécurité en eau" et "distance à l'objectif", peuvent être synthétisés en un critère représentant la capacité de la stratégie à satisfaire la consommation prévisionnelle, d'une part au cours du déroulement de la journée, d'autre part en ne puisant pas sur les réserves. Ce critère synthétique de premier niveau sera appelé plus loin "quantité".

De même, les critères "pression maximale" et "nombre de changements de configuration", sont regroupables en un critère nommé "protection", qui intègre les aspects liés aux risques de rupture du réseau, et la fatigue des équipements (pompes, vannes).

Les critères "coûts", et "sécurité-contrat-EDF", ont été rassemblés en un critère agrégeant le coût économique et le risque de paiement d'une pénalité.

A ce premier niveau, le critère "régime nourricier", reste seul. D'ailleurs, ce critère ne concerne pas les stratégies partielles, mais seulement les stratégies complètes. Pour la sélection à l'intérieur d'une plage horaire, ce critère n'est donc pas pris en compte.

b) Les regroupements de second niveau

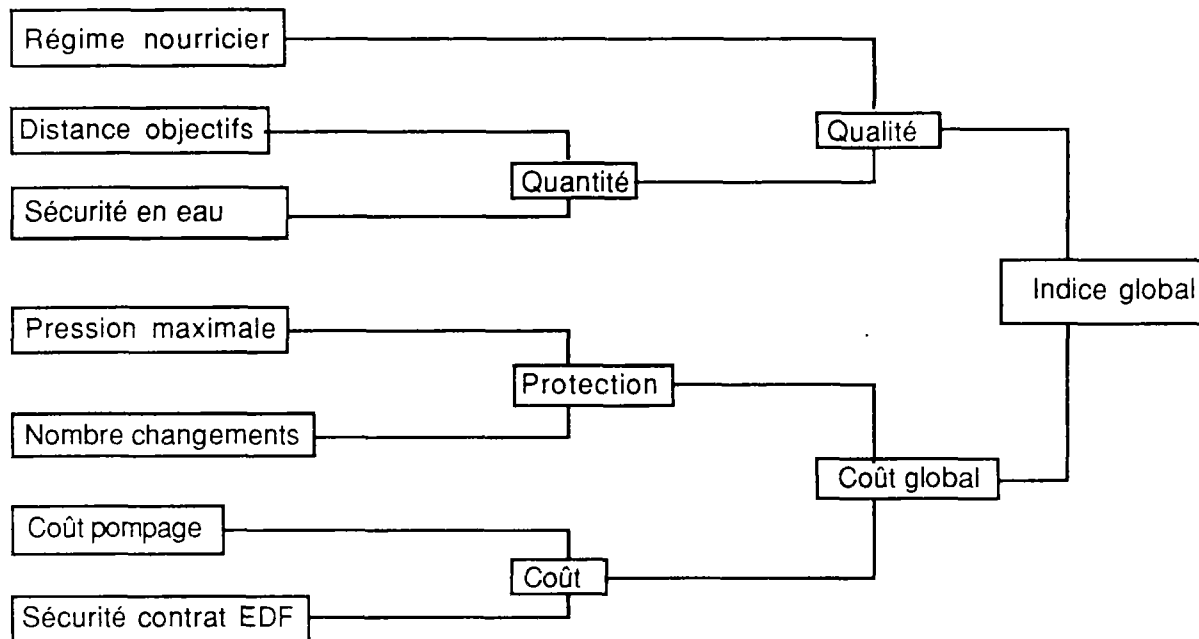
Pour les stratégies complètes, il est apparu pertinent de regrouper le critère élémentaire "régime nourricier" avec le critère "quantité" dans un niveau donnant une indication globale sur la production, intégrant les aspects quantité et qualité, même si la régularité du régime nourricier n'est qu'un élément de contribution à la qualité de l'eau produite.

Pour toutes les stratégies, les critères "coût global et protection", sont agrégés en un seul indicateur économique qui assemble le coût immédiat, et les risques de coût induits, par rupture de canalisation, défaillance de pompes, ou dépassement de contrat.

c) Le regroupement de troisième niveau

L'indice de satisfaction de la demande et l'indicateur économique sont finalement rassemblés en un seul indice caractérisant globalement la stratégie.

d) Le schéma du regroupement



3) La définition des classes de qualification du critère global

Le problème est de croiser les indications données pour 2 critères, chacune des stratégies étant décrite par trois classes sur l'ensemble des critères.

Ceci revient concrètement à remplir un tableau à 9 cases, en indiquant dans chaque case, la résultante des classes respectives des 2 critères élémentaires. Quatre cas intéressants de tableau résultant ont été définis.

a) La déqualification maximale

Dans ce cas de combinaison, représentée dans le tableau ci dessous, c'est l'indication la plus mauvaise d'un critère qui donne la classe de qualification du critère global. Le terme de déqualification maximale vient du fait que le critère résultant est bon dans un seul cas, moyen dans 3 cas, et mauvais dans 5 cas.

-----	*-----*	*-----*	*-----*	*-----*
* Critère 1	* bon	* Critère 2	* moy	* mau
* bon	* bon	* moy	* mau	* mau
* moy	* moy	* moy	* mau	* mau
* mau	* mau	* mau	* mau	* mau
-----	*-----*	*-----*	*-----*	*-----*

b) La déqualification minimale

Dans ce cas de combinaison, représentée dans le tableau ci dessous, c'est l'indication la meilleure d'un critère qui donne la classe de qualification du critère global. Les nombres respectifs de critères résultants "bon, moyen, mauvais" sont inversés par rapport au cas précédent.

-----	*-----*	*-----*	*-----*	*-----*
* Critère 1	* bon	* Critère 2	* moy	* mau
* bon	* bon	* bon	* bon	* bon
* moy	* bon	* moy	* moy	* moy
* mau	* bon	* moy	* mau	* mau
-----	*-----*	*-----*	*-----*	*-----*

c) La déqualification fortement pondérée

Dans ce cas de combinaison, représentée dans le tableau ci dessous, c'est l'indication la meilleure d'un critère qui donne la classe de qualification du critère global.

*-----	*-----	*-----	*-----	*
*	*	Critère 2		
* Critère 1	* bon	* moy	* mau	*
*	*-----	*-----	*-----	*
* bon	* bon	* bon	* moy	*
*	*	*	*	*
* moy	* moy	* moy	* mau	*
*	*	*	*	*
* mau	* mau	* mau	* mau	*
*	*-----	*-----	*-----	*

d) La déqualification faiblement pondérée

Dans ce cas de combinaison, représentée dans le tableau ci dessous, c'est l'indication la meilleure d'un critère qui donne la classe de qualification du critère global.

*-----	*-----	*-----	*-----	*
*	*	Critère 2		
* Critère 1	* bon	* moy	* mau	*
*	*-----	*-----	*-----	*
* bon	* bon	* bon	* bon	*
*	*	*	*	*
* moy	* moy	* moy	* mau	*
*	*	*	*	*
* mau	* mau	* mau	* mau	*
*	*-----	*-----	*-----	*

4) Les classes dans le cas d'agrégation pour les stratégies

Tout l'expertise n'a pas pu être discutée à fond, et les agrégations qui sont proposées ici ne sont représentatives que de certaines opinions, appuyée sur les résultats des questionnaires.

a) Les agrégations du premier niveau

Pour le critère quantité, le type d'agrégation retenue est une pondération forte par rapport au critère sécurité, qui a été défini comme le

plus important lors de l'enquête faite par questionnaire auprès des chefs de poste.

Pour le critère protection, présentant des risques de même nature, ou perçus comme tels par les chefs de poste, c'est une déqualification maximale qui pourrait être envisagée.

Pour le critère coût, l'aspect risque de dépassement a été classé second dans la hiérarchie définie par les chefs de poste. Ce point oriente vers une pondération forte par rapport à ce critère.

b) Les agrégations de deuxième et troisième niveau

Les études de cas et les entretiens qui les ont suivi, n'ont pas permis d'aborder ce niveau, mais ceci constitue une voie à explorer.

5) Méthodologie de l'utilisation de l'agrégation

Tout d'abord, à chaque stratégie encore sélectionnée à ce niveau, est affecté un indice global, en utilisant l'arbre de regroupement, avec les règles d'agrégation qui ont été définies. Les comparaisons se font à un niveau synthétique très élevé, donc facile à gérer, puisqu'il ne prend en compte qu'un seul paramètre, la classe de la stratégie par rapport au critère global.

Ceci correspond à la phase de recherche d'une information synthétique

L'ensemble des stratégies candidates est alors filtré par ce critère global, et ne sont retenues que les stratégies présentes dans la meilleure classe, c'est à dire ayant la qualification "bon", ou si cette classe est vide, ayant la qualification "moyen", ou encore, par défaut, la qualification "mauvais". On peut d'ailleurs remarquer que dans ce dernier cas, le filtre "critère global" est inefficace.

Ce filtre réduit le nombre de stratégies candidates, mais il est encore peu fréquent que ce nombre soit égal à 1.

La logique de recherche d'informations complémentaires fait que l'on va décomposer le critère global, en niveau synthétique immédiatement inférieur, pour rechercher des différences entre stratégies sur ce niveau plus fin.

Les stratégies encore candidates sont donc comparées sur 2 critères, mais cette comparaison plus lourde à gérer ne concerne qu'un nombre réduit de stratégies. On retrouve donc bien là, l'idée directrice de la

démarche, qui veut que la recherche des compléments d'informations ne soit effectuée que lorsqu'elle est pertinente.

6) Intérêt de l'agrégation de critères

Les techniques d'agrégation de critères, c'est à dire la construction de l'arbre à partir des critères élémentaires, et la définition des classes du critère agrégé peuvent être modulées de façon simple et explicite en fonction du contexte.

En effet, l'agrégation permet de mettre facilement en relation les priorités fixées pour sélectionner une stratégie en fonction du contexte qui a été identifié.

D) Une évaluation des gains potentiels

La phase d'évaluation a permis de déterminer les écarts maximaux de coûts entre une bonne stratégie et une autre qui est fortement déqualifiée. Cet écart est de l'ordre de 0,5 centime par m³ élevé, ce qui représente au plus 2000 F par jour. Mais il n'est pas évident, compte-tenu de l'ensemble des paramètres non maîtrisés, en particulier concernant la prévision, que les chefs de poste choisissent systématiquement une mauvaise stratégie, et PILOTE la meilleure.

Par contre, en situation normale, PILOTE produit des stratégies cohérentes, qui pourraient être affinées par des améliorations sur le modèle de simulation.

A partir de ces constatations, l'argumentation sur le développement de PILOTE, en tant que projet, change d'optique. Le système, conçu pour l'aide à la décision, pourrait constituer le noyau central d'un système de conduite automatique, adapté à un ensemble de situations répertoriées. Le pilotage serait repris manuellement dès que l'on sort d'un cas strictement défini dans le cadre de l'automatisme.

Mais, du point de vue méthodologique, le plus important est que l'utilisation quotidienne de PILOTE, et surtout l'analyse quotidienne des erreurs, se révèle une source importante de renforcement de la connaissance, en particulier pour l'ajustement des seuils et objectifs définis dans la base de connaissance.

Ceci a orienté la réflexion vers un deuxième axe : l'apprentissage automatique.

V) UNE ORIENTATION VERS LES PROGRAMMES D'APPRENTISSAGE

A) Une réflexion à mener

Le suivi des développements informatiques en intelligence artificielle montre qu'un volet apprentissage est maintenant intégré dès la phase de définition des projets.

Les concepteurs de projet raisonnent en termes d'apprentissage automatique pour ne pas perdre l'expertise générée par la mise au point et l'utilisation du logiciel en exploitation courante, ce qui est le cas de figure du projet PILOTE.

L'apprentissage opérationnel d'un système implanté sur une usine n'est pas encore à l'ordre du jour, par contre, il est envisageable de mener une réflexion méthodologique prospective orientée vers l'apprentissage automatique.

B) Points pouvant faire l'objet d'un apprentissage dans PILOTE

Plusieurs points issus de l'expertise initiale sont sujets à une remise en cause ou des ajustements au vu des résultats des premiers tests:

- les objectifs, et les seuils ou fourchettes d'élimination,
- les classes d'identification des contextes,
- les processus de traitement de l'information.

1) L'ajustement des objectifs et des seuils

Il a été expliqué dans le chapitre 3 comment ont été définis les seuils et les objectifs, qui représentent des valeurs consensuelles. Ces valeurs ont servi de référence, mais il n'a jamais été démontré qu'elles sont optimales. Une analyse critique a posteriori des décisions peut servir à ajuster ces valeurs en simulant différents cas de figure sur des séries de journées.

Pour l'ensemble des conclusions qu'ils formulent, les systèmes experts retracent aujourd'hui l'enchaînement des règles utilisées. Il est donc possible, dans le cas où une décision erronée a été prise par la base de règles, de retrouver les règles qui l'ont induite.

A partir de là, il est tout à fait envisageable de constituer un fichier des décisions erronées prises par le système.

Ce fichier pourrait alors être examiné à intervalles réguliers pour modifier les règles incriminées. Ces modifications seraient faites en

simulant à postériori la prise de décision, mais en changeant à tour de rôle un paramètre, par exemple :

- modification d'un objectif de fin de période de planification,
- modification d'un seuil d'élimination,
- modification du mode d'utilisation des critères.

2) L'identification des contextes

Les contextes ont un rôle extrêmement important dans la démarche, parce qu'ils définissent les grandes lignes du raisonnement et activent les schémas décisionnels. La constitution au fil des jours d'une bibliothèque de cas est un point de départ pour stabiliser et affiner le répertoire des situations que le système a à traiter.

Les profils types de consommation en eau du réseau, d'évolution de pression de refoulement, de consommation électrique dans l'usine, ont été, dans un premier temps, établies à partir de données historiques.

Mais au fur et à mesure que le système fonctionnera, il enregistrera de nouvelles données. Les informations acquises au cours du fonctionnement pourront contribuer à affiner les courbes types existantes ou à créer de nouvelles références.

Ce type d'apprentissage, qui peut être compris comme un auto-ajustement du programme en fonction des données qu'il acquiert, ne présente pas de difficultés particulières, et la méthode développée à Maisons-Laffitte dans le cas du respiromètre pourrait être réutilisée ici [LANNUZEL, JESTIN 87].

3) Le renforcement du processus de traitement de l'information

L'utilisation du système sur un grand nombre de cas caractérisera la pertinence des différentes informations manipulées. La recherche d'indicateurs de sélection de stratégies plus efficaces ou synthétiques, par une réorganisation de la connaissance, constitue un champ d'investigation pour l'apprentissage.

Ces 3 points demandent bien sûr à être approfondis et mis en relation avec des techniques d'apprentissage, qui sont présentées plus loin.

C) Éléments caractérisant les programmes d'apprentissage

1) Définition de l'apprentissage

Plusieurs définitions du terme apprentissage existent aujourd'hui et recouvrent des concepts parfois différents.

D'un point de vue commun, un système apprend s'il opère en lui-même les changements qui améliorent la réalisation d'une tâche déterminée.

M SIMON, un des premiers chercheurs ayant travaillé sur le sujet, a défini dans la presse l'apprentissage comme "n'importe quel processus par lequel un système améliore sa performance".

Selon d'autres sources, plus proches de ceux qui réalisent actuellement des programmes, l'apprentissage est compris comme une acquisition de connaissances explicites ou de schémas de raisonnement, ou encore comme une nouvelle organisation de la connaissance dans des représentations plus efficaces.

S'il est difficile de donner une définition exhaustive de la notion d'apprentissage, il est par contre possible d'indiquer quelques points de repère sur les programmes d'apprentissage automatique.

Un programme d'apprentissage automatique est caractérisé par son champ d'investigation et sa finalité. Il importe en effet de savoir ce que le système apprend, et quelle sera l'utilisation ultérieure des connaissances acquises. Les finalités les plus courantes de ces programmes sont présentées dans le paragraphe 2 .

Tout apprentissage suppose une méthodologie, et donc la compréhension des différentes étapes. Les grandes lignes des principales méthodologies sont données dans le paragraphe E, avant d'être détaillées dans l'annexe 2, lors de la description des programmes.

2) Finalité des principaux programmes d'apprentissage

De manière générale, la tâche d'apprentissage consiste à enrichir la connaissance d'un système à partir d'un ensemble d'informations qui lui sont fournies, ce qui peut prendre différents aspects dont les principaux sont donnés ci-dessous.

a) La définition de concepts :

Cette tâche consiste à décrire un concept, dans un langage de représentation donné, à partir d'un jeu d'exemples et de contre-exemples présentés au programme. Le programme est ensuite capable d'identifier l'appartenance d'objets nouveaux au concept décrit. Cette tâche est parfois opérationnalisée par la construction d'un arbre de décision.

b) La structuration de données :

Dans ce type d'apprentissage, le programme cherche à regrouper les objets en fonction de leur similitude. Ces regroupements peuvent avoir plusieurs niveaux hiérarchiques et conduire à la construction d'une taxinomie.

c) L'amélioration de schémas de raisonnement :

Un certain nombre de problèmes sont résolus par l'application d'une séquence d'opérateurs transformant l'état initial du problème en but. Certains programmes d'apprentissage précisent les conditions d'utilisation des différents opérateurs, et proposent des séquences complètes correspondant à des problèmes types.

d) La réorganisation des connaissances :

Certains programmes n'ajoutent aucune connaissance sur le domaine dans lequel ils opèrent, mais ont pour objet d'améliorer la recherche et l'exploitation des informations dans une base de connaissances.

e) La modifications de valeurs de paramètres :

Ces programmes modifient des valeurs de paramètres d'une fonction pour la caler par retro-action.

f) Création ou modification de règles de production :

Cet apprentissage permet d'augmenter la quantité d'informations déductibles de l'ensemble initial des faits, par création de nouvelles règles, ou la précision de ces informations, en modifiant les conditions d'application ou les actions des règles existantes.

g) L'importance de l'induction dans l'apprentissage :

La description sommaire de quelques programmes indique que la notion d'apprentissage est souvent liée à celle de raisonnement inductif, aussi ce point va t'il être développé dans le paragraphe suivant.

3) Raisonnement inductif et déductif

a) La déduction :

Le raisonnement déductif consiste à produire des faits nouveaux F_n à partir de faits initiaux F_i et d'une théorie ou ensemble de règles E_r .

$$F_i \text{ et } E_r \Rightarrow F_n$$

b) L'induction :

b.1) Le paradigme inductif :

Etant donnés un ensemble d'observations ou de faits F et une connaissance de base préalable CB , il s'agit de trouver une assertion ou hypothèse H , qui, combinée avec la connaissance de base CB , implique F l'ensemble des faits et observations dans l'état de connaissance du système.

$$H \text{ et } CB \Rightarrow F$$

b.2) La formulation de théories et le principe de réfutation

Toute induction conduit à la formulation d'une nouvelle théorie. Par exemple, la généralisation des 2 faits suivants :

GREC(socrate)&PHILOSOPHE(socrate)
GREC(platon)&PHILOSOPHE(platon)

conduit à définir les 2 théories (formées d'une seule hypothèse) :

GREC(X) => PHILOSOPHE(X)	(tous les Grecs sont des philosophes)
PHILOSOPHE(X) => GREC(x)	(tous les philosophes sont Grecs).

L'introduction d'un 3ème fait "ALLEMAND(kant)&PHILOSOPHE(kant)" va invalider la 2ème théorie.

Toute théorie obtenue par induction est considérée valable dans un domaine tant qu'elle n'a pas été réfutée. Le principe de réfutation va également servir à tester le fondement et l'utilité des théories envisageables, les théories difficilement réfutables, parce qu'elles font appel à des faits qu'on ne peut obtenir, sont considérées comme peu utiles.

b.3) Caractéristiques des hypothèses :

Généralement, un raisonnement inductif conduit à la définition d'un jeu d'hypothèses qui sont considérées comme vraies jusqu'à ce qu'elles soient contredites par la production d'un fait ou une observation nouvelle.

Toutes les hypothèses n'ont pas le même degré de pertinence, aussi sont elles analysées selon trois dimensions : la simplicité, l'efficacité et le niveau d'abstraction. Suivant la nature de l'apprentissage, le programme hiérarchise les hypothèses selon un critère de préférence prenant en compte ces trois dimensions selon une pondération définie le plus souvent à priori.

4) Eléments d'un programme d'apprentissage automatique

Les programmes d'apprentissage viennent, dans la plupart des cas, se greffer sur un système informatique existant, dont l'élément central est le module de réalisation. C'est le fonctionnement de ce module que le programme d'apprentissage cherche à optimiser. Ce module est parfois formé d'une seule fonction avec ses paramètres; dans d'autres cas il est constitué par un ensemble de règles articulées entre elles. Son architecture et sa complexité vont largement déterminer la méthodologie de l'apprentissage.

Le schéma ci-dessous représente les différents éléments caractéristiques d'un programme d'apprentissage inductif.

Base initiale de connaissances

MODULE DE REALISATION

Utilisation par le système
des hypothèses nouvelles
Production de résultats

Base de connaissances enrichie

Environnement

fournissant des informations

MODULE D'APPRENTISSAGE

Traitement des informations
Génération d'hypothèses

Critique des résultats
Validation, rejet ou création
d'hypothèses

Une session d'apprentissage commence par la prise des informations fournies par l'environnement. La nature de ces informations et leur présentation dépendent de la tâche d'apprentissage demandée et du mode de structuration de la connaissance dans la base.

a) L'architecture de la base de connaissances :

Dans les programmes d'apprentissage, la connaissance est structurée selon deux formes principales : les vecteurs de caractéristiques ou le calcul des prédicats.

Les vecteurs de caractéristiques décrivent les objets avec un ensemble fixé de propriétés (descripteurs, attributs) auxquelles sont attribuées un ensemble fini de valeurs. Par exemple :

objet.couleur = vert; objet.longueur = 10.

Les prédicats décrivent à la fois les propriétés des objets structurés et leurs relations.

ROUGE(X)&SUR(X,Y)

La comparaison de deux descriptions, fréquemment utilisée en apprentissage, est plus facile avec les vecteurs, mais la manipulation d'objets est limitée par rapport aux possibilités du calcul des prédicats.

b) Le module d'apprentissage :

Un module d'apprentissage est généralement composé de deux parties ayant des rôles différents : le générateur d'hypothèses et le critique. Le générateur d'hypothèses commence par définir, tout en le délimitant l'espace de recherche. Suivant la nature de l'apprentissage, cet espace est composé d'un ensemble de descriptions plausibles, d'intervalles de valeurs pour jeux de paramètres, de schémas de décisions, etc...

Ensuite ce module génère des hypothèses sur les descriptions, valeurs de paramètres, et ensuite hiérarchise ces hypothèses selon des critères de préférence.

Les hypothèses sont alors testées par le module de réalisation.

C'est à ce niveau qu'intervient le second module caractéristique des raisonnements inductifs : le critique, dont la fonction est d'analyser les conséquences des hypothèses en cours d'évaluation, de les valider, les rejeter ou les modifier avant une nouvelle évaluation.

La base de connaissances n'est enrichie qu'après une validation lors de l'étape de critique des résultats.

c) Le module de réalisation

Il est le point central du système d'apprentissage, car ce sont ses actions que le module d'apprentissage essaie d'améliorer. L'efficacité et la facilité de l'apprentissage vont dépendre de la complexité, de la modifiabilité et de la transparence de ce module.

Le problème de l'intégration ou de la modification d'une connaissance dans un ensemble existant n'est pas trivial, car le système doit considérer les interactions possibles entre les nouvelles connaissances et les anciennes.

Plus ce module est complexe, plus les interactions potentielles sont nombreuses, et le système doit avoir un moyen d'évaluer les hypothèses proposées par le module d'apprentissage.

D'autre part, pour renforcer ou rejeter des éléments de connaissances, le système doit avoir accès aux actions internes du module de réalisation, ce qui suppose que ce module garde la trace des raisonnements.

d) L'apprentissage : l'exploration d'un espace de recherche :

D'une manière générale, les programmes d'apprentissage explorent un espace de recherche en étant guidés par une connaissance de base, liée au domaine et au type d'apprentissage.

5) Classification des programmes d'apprentissage

Un très grand nombre de programmes d'apprentissage automatique ont été développés ou sont en cours de développement aujourd'hui.

Tout savoir étant d'abord classificatoire, un premier travail consiste à regrouper ces programmes par grandes familles pour identifier leurs éléments caractéristiques et principaux.

La finalité de la tâche d'apprentissage, le champ d'application, ou la méthodologie suivie sont autant de clefs sur lesquels peuvent s'appuyer des classifications, qui ont été présentées dans les ouvrages de référence comme MACHINE LEARNING.

Les classifications phénoménologiques souvent proposées sont basées sur le type principal des inférences utilisées et sur la finalité de l'apprentissage. Les classes sont indexées sous la forme "apprentissage par ..." (instruction, analogie).

BUNDY, de l'université d'Edimbourg, a proposé une classification analytique basée sur les algorithmes des programmes de gestion des hypothèses.

D'autres classifications prennent appui sur le domaine de la base de connaissance que le programme cherche à améliorer.

Ce rapport présente une approche inspirée de la classification phénoménologique décrite dans le "HANDBOOK OF ARTIFICIAL INTELLIGENCE", au chapitre sur l'apprentissage. Cette classification présente l'inconvénient de distinguer des types d'apprentissage dont les algorithmes sont très proches, lorsqu'ils n'ont pas la même finalité. Mais, comme une classification analytique est difficilement lisible, la classification phénoménologique a été préférée parce qu'elle est plus proche des modèles comportementaux humains.

La taxonomie, présentée à la page suivante, est basée sur la nature et le degré d'abstraction de l'information fournie au programme.

Cette classification de base recouvre en fait deux types de programmes nettement différents. La première classe est constituée par les programmes qui ne créent pas de connaissances nouvelles dans le domaine, mais seulement restructurent la connaissance initiale pour la rendre plus opérationnelle.

Une seconde classe est constituée par les programmes qui génèrent des hypothèses, dont une partie viendra, après la phase de validation, enrichir la base de connaissances initiale.

TYPE	NIVEAU D'ABSTRACTION DES INFORMATIONS	DESCRIPTION DE LA TACHE D'APPRENTISSAGE	SYSTEME ENRICHI
implantation	informations utilisables sans traitement	organisation du stockage des informations	mémorisation de situations pour éviter de recalculer
avis ou instructions	niveau élevé	déterminer les détails non fournis adapter les concepts	concepts adaptés aux situations courantes de la base de connaissance
à partir d'exemples	informations nombreuses spécifiques et détaillées	définir des concepts regrouper, classer	descriptions de concepts pour classer les objets nouveaux
à partir d'observations	informations nombreuses non structurées	rechercher les similitudes regrouper les objets	nouveaux concepts taxinomies
par expéri- mentation	cas générés par le système lui-même	définir des plans, actions conditions d'application d'opérateurs	schémas de résolution heuristiques pour le choix d'opérateurs
par recherche d'explications	théorie du domaine cas résolu	trouver la trace de la preuve, définir un ensemble de descriptions	descriptions structu- relles
par analogie	concept de niveau élevé adapté à une tâche distincte	découvrir l'analogie, sa portée, et compléter les concepts ou les objets de la base de connaissance.	concepts, attributs nouveaux relations nouvelles entre attributs.

Dans le premier cas, appelé apprentissage par implantation, le programme reçoit des informations exactement du niveau dont il a besoin. Aucune hypothèse n'est à formuler et l'information reçue n'a pas à être traitée ou interprétée avant d'être utilisée.

Dans les cas d'apprentissage par instruction, l'environnement fournit des informations abstraites ou générales et le programme doit alors faire des hypothèses sur les détails non présents pour adapter les informations reçues au contexte particulier du problème traité.

A l'opposé, les informations fournies au système peuvent être trop spécifiques et détaillées, comme par exemple une liste d'exemples ou d'observations, et le travail d'apprentissage consiste alors à extraire des règles d'un niveau plus général pouvant s'appliquer ensuite, pour classer de nouveaux objets.

Les différents types d'apprentissage présentés dans ce tableau, sont détaillés en annexe.

VI) CONCLUSION DE LA PHASE D'EVALUATION

De cette phase d'évaluation, on peut retenir synthétiquement les points suivants.

En ce qui concerne le modèle prévisionnel, l'écart entre prévisions et réalisations est resté inférieur à 5 %, dans un peu plus des 2/3 cas, sur une période de 5 mois. Comme, en deça de cet écart, les décisions ne sont pas sensiblement affectées, on peut conclure que le modèle est acceptable, dans de développement actuel du projet.

Le modèle de simulation pose par contre un problème préoccupant pour la suite du projet. Les dérives observées sont importantes, et faussent les décisions. Il y a donc là, un travail de réactualisation des courbes de pompe, et de recalage de modèle, qui est une condition préalable à tout développement ultérieur.

L'évaluation des heuristiques marque à la fois l'intérêt et les limites de l'intégration de l'expérience acquise dans la conduite hydraulique d'une usine. Les heuristiques permettent bien de restreindre l'exploration sur les stratégies les plus intéressantes, à l'intérieur des périodes de planification. Mais, les heuristiques deviennent insuffisantes dès lors que l'on raisonne à l'échelle de la journée. Cela est apparu évident, lors des comparaisons des décisions de PILOTE avec les décisions humaines. En conséquence, l'optique et la finalité du système changent. Le renforcement des heuristiques passe par l'intégration de l'expérience acquise en fonctionnement. C'est bien en ce sens que les programmes d'apprentissage automatique pourraient apporter des réponses dans cette voie.

En ce qui concerne l'amélioration des heuristiques, la priorité est de renforcer l'expertise au niveau des modes d'utilisation des critères et de leurs classes de qualification. L'agrégation de critères reproduit mieux qu'une sélection lexicographique, la démarche suivie par les chefs de poste. Cette constatation indique la voie à suivre, mais il reste à définir tous les niveaux de regroupements, et les classes de qualifications qui en résultent.

Le passage d'une maquette à un prototype implique donc, à court terme, les trois priorités suivantes :

- une prise en compte rapide de l'aspect méthodologique sur le contrôle d'exécution,
- un renforcement du modèle de simulation,
- un affinage des procédures de choix des stratégies.

Pour un développement à plus long terme, les besoins ne sont plus les mêmes. Certaines heuristiques ont fixé des valeurs de référence, comme les objectifs de remplissage, et les seuils d'élimination. Ces valeurs ont été recueillies auprès des experts, et après la phase de questionnaires, représentent un consensus, limité à un nombre de situations répertoriées. Indispensables pour démarrer le projet, elles ne doivent pas être considérées comme des références immuables. En particulier, le suivi régulier du fonctionnement du système, et une analyse systématique à posteriori des décisions, conduiront à une réévaluation, voire une remise en cause de ces valeurs initiales.

C'est sur ce point, que les programmes d'apprentissage auront un apport méthodologique important.

Ainsi, cette thèse a recueilli et intégré les heuristiques, qui ont contribué à l'initialisation du projet. L'évaluation a mis en évidence tout l'intérêt de cette démarche, mais aussi la nécessité de ne pas considérer la connaissance initiale comme immuable. Les jeux d'heuristiques peuvent continuer à être enrichis par l'expérience que le système acquiert lui-même en fonctionnant. La définition d'une démarche d'intégration des nouveaux résultats constitue alors une nouvelle voie à explorer.

BIBLIOGRAPHIE

- BARR A 1982
FEIGENBAUM E. Handbook of artificial intelligence. Tomes 1 et 2
Ed Kaufman
- BLACK W.J. 1988 Systèmes intelligents basés sur la connaissance.
Ed Masson.
- BONNET A. 1984 L'intelligence artificielle, promesses et réalités.
Ed Interéditions
- BONNET A. 1986
HATON J.P. Systèmes experts. Vers la maîtrise technique.
Ed Interéditions
- TRUONG NGOC J.P.
- BOOSE J.H. 1986 Expertise Transfert for Expert system design.
Ed Elsevier.
- BORDAGE G 1980 Representation of medical knowledge categories and
prototypes. Communication AAMC RIME
Washington, D. C., Octobre 80.
- BRIAND R. 1988 Méthodes de développement de systèmes experts
Ed Eyrolles.
- BUCHANAN B. 1978 DENDRAL et META-DENDRAL, their application
FEIGENBAUM E. dimension. In Artificial Intelligence, vol 11, mai 78
- CARBONNEL J. 1984 Machine learning, an artificial intelligence approach
MICHALSKI R. Tomes 1, 2, 3.
MITCHELL T. Ed Morgan Kaufman Publishers.
- CHARNIAK E. 1984 Introduction to artificial intelligence.
MAC DERMOT D. Ed Addison-Wesley.
- COHEN P.R. 1985 Heuristic reasoning about uncertainty : an artificial
intelligence approach.
Pitman Advanced Publishing Program
- COHEN P.R. 1985 Handbook of artificial intelligence. Tome 3.
FEIGENBAUM E. Ed Kaufman.

- COT J.P. 1986 Formes de représentations de la connaissance.
In Le Monde Informatique du 22 septembre 86.
- DE GROOT A 1965 Thought and choice in chess. Ed Mouton.
- FARRENY H. 1987 Eléments d'intelligence artificielle. GHALLAB M.
Ed Hermes.
- FARRENY H. 1985 Systèmes experts, principes et exemples.
Ed Cepadues.
- FIESCHI M. 1986 Intelligence artificielle en médecine. Ed Masson
- GALLOUIN J.F. 1988 Transferts de connaissances pour systèmes experts
Techniques et méthodes. Ed Eyrolles.
- GONDRAN M. 1984 Introduction aux systèmes experts. Ed Eyrolles
- HART A. 1988 Acquisition du savoir pour les systèmes experts.
Editions Masson.
- HAYES-ROTH F 1983 Building expert systems
LENAT D.B. Addison-Wesley Publishing Company
WATERMAN D.A.
- HOLSAPPLE G. 1986 Expert system using GURU.
WHINSTON Ed DOW JONES IRWIN.
- ISE-CEGOS 1986 Guide d'apprentissage de GURU.
- KAUFMANN A. 1987 Nouvelles logiques pour l'intelligence artificielle
Ed Hermes.
- KODRATOFF Y. 1986 Leçons d'apprentissage symbolique automatique
Cepadues Editions.
- LANNUZEL 1987 Biosurveyor, a knowledge based system for
JESTIN operators of waste water treatment plants.
COUSIN, In proc. IV, International conference in urban
MEYNY storm drainage, Lausanne., sept 87.

- LANNUZEL P.1988 Comparaison de modèles prévisionnels.
PARENTE E. Rapport remis à la C.G.E.
- LAURIERE J.L.1987 Intelligence artificielle, Résolution de problème par
l'homme et la machine. Ed Eyrolles
- LAURIERE J.L.1988 Intelligence artificielle, Représentation des
connaissances. Ed Eyrolles.
- LENAT D. B 1982 The nature of heuristics. Artificial intelligence n°19.
- LEQUENNE P. 1987 Outils d'aide à la décision pour la production d'eau
potable. Rapport remis à la C.G.E.
- NEURON DATA 1987 Reference manual of Nexpert Object.
- NEWELL 1969 GPS, a program that simulates human thought.
SIMON Ed Mc Graw-Hill.
- NEWELL 1972 Human problem solving.
SIMON Ed Englewood cliff, NJ, Prentice-Hall.
- PAUKER 1977 Analyzing and simulating. Taking the history of the
present illness context formation.
Ed SNEIDER SAGWALL HEIN.
- PEARL 1984 Heuristics : Intelligent search strategies for
computer problem solving. Ed Addison Wesley.
- PENET G 1987 Etude de la consommation d'eau de l'usine de
Neuilly. Rapport remis à la C.G.E.
- POPLE 1982 Heuristics Methods for imposing structure on
ill structured problems. In Artificial intelligence
in medecine. Ed P SZOCOVITS.
- QUEINNEC A. 1984 LISP, Mode d'emploi. Ed Eyrolles.
- RICH E. 1983 Artificial intelligence. Ed Mac-Graw Hill.
- ROY B. 1985 Méthodes multicritères d'aide à la décision.
Ed Economica.
- SHAPIRO E. 1986 Art of PROLOG
STERLING L MIT Press.

- SHAPIRO E. 1987 Encyclopedia of artificial intelligence. Tomes 1 et 2.
STUART C. Ed Wiley Interscience.
- SHORTLIFFE E.H. Computer based medical consultation : Mycin.
1976 Ed Elsevier, New York.
- TURNER 1986 Logiques pour l'intelligence artificielle. Ed Masson
- VOGEL C. 1988 Génie cognitif. Edition Masson.
- VOOGD H.1983 Multicriteria evaluation for urban and regional
planning. Pion limited, London.
- WATERMAN D.A. A guide to expert systems. Ed Addison-Wesley
1985
- WINSTON P.H. 1981 Artificial intelligence. Ed Addison Wesley.

ANNEXE 1

Δ Δ Δ

Caractéristiques de quelques outils de développement de systèmes experts.

O O O

SOMMAIRE DE L'ANNEXE

A) Les langages de manipulation symbolique	3
1) Un langage général : LISP	
2) Un langage spécialisé : PROLOG	
B) Un générateur dans un logiciel intégré : GURU	10
1) La notion de logiciel intégré	
2) La structuration de la connaissance	
3) Le contrôle du raisonnement dans GURU	
C) Un générateur basé sur la gestion d'objets : NEXPERT-Objet	13
1) Les concepts de base de Nexpert	
2) Le contrôle dans Nexpert	
D) Un générateur basé sur la logique des prédicats d'ordre 1 : l'Expert-Kit	
1) Les concepts de base	17
2) La représentation de la connaissance	
3) Les stratégies de contrôle	
E) Un générateur basé sur la logique des prédicats d'ordre 2 : SNARK	
1) Les concepts de base	19
2) La représentation de la connaissance	
3) Les stratégies de contrôle	

A/ Les langages de manipulation symbolique

Il est possible de construire un système-expert à partir de n'importe quel langage informatique, comme le FORTRAN, le PASCAL, ou le C. D'ailleurs, la plupart des outils générateurs de système-expert sont écrits dans l'un de ces langages. Deux langages de manipulation symbolique servent de référence en intelligence artificielle : LISP et PROLOG.

1/ Un langage général : LISP

a/ Un langage historique de traitement de liste :

Le langage LISP a été créé en 1958 par Mac Carthy et conçu dès son origine pour travailler sur des listes, c'est-à-dire des collections d'items entourés de parenthèses où chaque item est soit un symbole, soit une autre liste.

LISP (LISt Processor) tient son nom de sa capacité à traiter les structures de liste, et est aujourd'hui le langage le plus utilisé dans l'intelligence artificielle parce qu'il présente les avantages suivants :

- une syntaxe simple,
- une gestion automatique de la mémoire,
- un traitement identique pour le code programme et les données,
- la puissance de la récursivité,
- les facilités d'extension et de modification.

b/ LISP : un langage interprété

LISP est un langage interprété, et l'interpréteur travaille par une boucle infinie de lecture-interprétation-écriture du résultat.

La méthode de programmation en LISP consiste à définir la fonction dont l'interprétation rend la solution au problème traité. Comme cette fonction est dans pratiquement tous les cas assez compliquée, elle fait appel à des fonctions auxiliaires, qui elles-mêmes peuvent faire appel à d'autres fonctions auxiliaires.

Le développement d'un programme en LISP consiste à décrire et définir l'ensemble des fonctions, puis à les assembler.

c/ La syntaxe de LISP : les atomes et les listes

Les atomes se répartissent en nombres et en symboles. Les nombres sont des entiers ou des réels, et les symboles des chaînes de caractères qui ne représentent pas des nombres. Les symboles sont considérés comme des variables, dont la valeur est cherchée dans chaque interprétation. Si une

variable n'a pas de valeur, le système renvoie en cours d'interprétation le message d'erreur "variable indéfinie".

Les listes se présentent comme des suites parenthésées, et représentent toujours une fonction appliquée à ses arguments. La fonction est le premier élément de la liste, appelé le CAR, et les arguments le reste de la liste (le CDR).

La liste vide "()" a pour valeur elle-même. La liste $(*, 2, 3)$ représente la fonction "*" appliquée à ses 2 arguments 2 et 3. LISP interprétera cette liste en rendant la valeur 6.

Dans le cas où le CAR de la liste n'est pas une fonction, et donc où la liste ne doit pas être interprétée, cette liste doit être précédée, du terme "QUOTE" ou d'une apostrophe.

d/ Les fonctions LISP

La définition des fonctions en LISP se fait à partir d'une fonction particulière, appelée "de, def, defun", selon le dialecte LISP.

Sa syntaxe est la suivante :

```
(def      nom de la fonction (paramètres)
          corps de la fonction)
```

Le carré d'un nombre x, fonction à un seul argument, est défini comme suit :

```
(def      carré (x)
          (* x x))
```

Une fois interprétée la fonction def avec ses arguments, la fonction "carré" peut être appelée seule, ou intégrée dans une autre fonction.

LISP, en interprétant (carré 4), rendra la valeur 16.

La fonction calculant le discriminant d'une équation du second degré $ax^2 + bx + c$, possède 3 arguments et se définit comme suit :

```
(def      disc (a b c)
          (- (carré b) (* 4 (* a c))))
```

Cette définition fait appel à la fonction carré, décrite précédemment. Une fois interprétée, cette nouvelle fonction def, la fonction disc peut être utilisée de façon quelconque, pourvu qu'elle ait 3 arguments. Ainsi :

```
(disc      2 3 1) sera interprétée et rendra la valeur -1.
```

Les différents langages LISP possèdent une bibliothèque de fonctions prédéfinies.

e/ Listes et traitement de listes

LIPS est un langage spécialisé dans la manipulation des listes et ces listes sont traitées en utilisant le caractère récursif dans la définition des fonctions, et sont décomposées en leur CAR d'une part et leur CDR de l'autre.

Cet aspect récursif qui permet à une fonction de s'appeler elle-même est illustré par le calcul d'une factorielle, et est opposé au calcul par itération.

Par itération, fac (n) est définie ainsi : (en Pascal)

i :=1; For i := 1 TO n DO i := i * i .

En utilisant la récursivité, fac (n) est définie ainsi :

**(def fac (n)
 (if (= 0 n) 1
 (* n (fac (- n 1)))))** **avec fac (0) = 1**

La récursivité permet d'écrire de façon condensée des fonctions parfois complexes, mais ceci demande la gestion d'une pile car toutes les valeurs intermédiaires du calcul sont gardées jusqu'à l'obtention du résultat final.

f/ La manipulation de faits et de règles

Les programmes LISP peuvent faire des déductions logiques à partir d'une base de données d'assertions représentées comme des listes LISP. Les assertions sont de deux sortes : la première sorte est constituée par des faits indiquant qu'un prédicat est vrai pour un objet déterminé. Par exemple l'assertion (**HOMME Socrate**) indique que le prédicat HOMME est vrai pour l'objet Socrate. L'autre sorte représente les règles indiquant qu'un prédicat en implique un autre. L'assertion (**TOUT HOMME MORTEL**) indique que le prédicat HOMME implique le prédicat MORTEL.

Construire un moteur d'inférences avec LISP consiste à définir les fonctions de manipulation de listes, ce qui est facilité par les caractères récursif et modulaire du langage.

Ces fonctions ne sont pas prédéfinies dans les langages de base, mais sont par contre commercialisées dans les sur-ensembles de LISP, tels que l'Experkit, décrit au chapitre suivant.

g/ Les machines LISP

Les machines LISP sont des calculateurs dédiés et conçus spécifiquement pour l'exécution de programmes LISP. La récursivité nécessite la gestion de piles, très gourmande en mémoire vive ; aussi ces calculateurs sont équipés de dispositifs microprogrammés spécifiquement pour la gestion de ces piles.

Mais avec le développement de la puissance de calcul des mini ordinateurs, l'intérêt de ces machines dédiées, beaucoup plus onéreuses, a diminué, et leur marché s'est considérablement restreint ces deux dernières années.

2/ Un langage spécialisé : PROLOG

a/ Un langage basé sur la logique des prédicats : PROLOG

PROLOG (PROgramming in LOGIC), développé à Marseille en 1972, par Alain COLMERAUER, est le plus connu des langages de programmation.

Il permet de représenter les connaissances (objets, relations entre objets, assertions, règles) directement dans les termes du calcul des prédicats restreint aux clauses de HORN, et de résoudre les problèmes posés dans le cadre de cet univers par des mécanismes de résolution et d'unification.

b/ La logique des prédicats restreinte aux clauses de HORN

Toute expression en logique des prédicats du premier ordre se transforme en une clause équivalente de la forme :

$$\{P1 \quad P2 \quad P3... \quad Pm\} \quad \Rightarrow \quad \{C1 \quad C2 \quad C3... \quad Cm\}$$

où $\{P1 \quad P2 \quad P3... \quad Pm\}$ représente un ensemble de prémisses (ou condition) qui doivent toutes être vraies, et $\{C1 \quad C2 \quad C3... \quad Cm\}$ un ensemble de conclusions dont au moins une est vraie.

Une expression sans condition correspond à une clause inconditionnelle ou assertion.

Les conditions et les conclusions sont des prédicats de la forme

$Pr(t1, t2... tm)$ où :

Pr est le nom du prédicat.

$t1... tn$ sont des termes qui peuvent être des constantes, des symboles, des variables ou des expressions.

La restriction aux clauses de HORN impose qu'il n'y ait qu'une seule conclusion dans une clause.

Exemples :

"plus-récent (a, b)" est une assertion comprenant 1 prédicat.

"plus-récent (a, c) <= plus-récent (a, b) & plus-récent (b,c)"
est une clause mettant en relation des prédicats à 2 arguments qui sont ici des variables.

"plus-brillant (x, y) ou plus puissant (x, y) <= plus récent (x, y)"
est une clause comprenant 2 conclusions que le langage PROLOG ne sait pas gérer.

c/ L'unification

Le mécanisme d'unification permet la mise en correspondance de deux expressions par un jeu de substitutions qui les rendent identiques.

Une substitution S est définie comme une application de l'ensemble des variables dans l'ensemble des termes.

E représente une expression, c'est-à-dire un terme ou un prédicat, et E_q une instantiation de E , c'est-à-dire une nouvelle expression identique à E dans laquelle les variables ont été remplacés par des constantes. Dans PROLOG, deux expressions E_1 et E_2 sont unifiables s'il existe une substitution S telle $E_1(S)$ soit identique à $E_2(S)$.

Par exemple, "**plus-récent (pompe-3, réservoir-2)**" n'est pas unifiable avec "**plus-récent (pompe-1, réservoir-3)**".

Les deux expressions "**plus-récent (pompe-3, x)**" et "**plus-récent (pompe-3, réservoir 2)**" sont unifiables avec la substitution " $x = \text{réservoir 2}$ ".

L'unification de deux expressions se fait par une comparaison des deux expressions, terme à terme, de la gauche vers la droite et en résolvant progressivement par substitution les désaccords.

d/ La méthode de résolution

Le principe de résolution de PROLOG est celle d'un démonstrateur de théorèmes, où le problème est posé sous la forme d'un prédicat, et où l'univers de connaissances est constitué par un ensemble de clauses.

PROLOG recherche toutes les solutions prouvant le prédicat ou prouvant que sa négation conduit à une contradiction, en utilisant le mécanisme d'unification et le backtrack (retour arrière).

Pour démontrer un prédicat P , PROLOG essaie d'unifier ce prédicat P avec les conclusions des clauses de l'univers des connaissances. Quand une unification est réussie, PROLOG tente de démontrer successivement les prémisses associées à la conclusion de la clause unifiée. L'évaluation de chaque prémisses se fait avec le même mécanisme.

Quand l'évaluation d'une prémisses échoue, il y a retour arrière vers la prémisses la plus proche où une autre unification est possible, et on reprend le mécanisme.

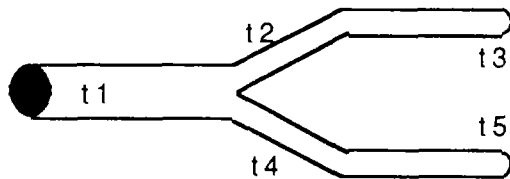
Lorsque toutes les prémisses correspondant aux différents sous-buts ont été prouvées, une solution est alors trouvée.

PROLOG continue en revenant au dernier point de choix, et reprend la recherche d'une autre solution. La recherche s'arrête quand toutes les solutions possibles ont été trouvées et l'ensemble de l'univers des connaissances exploré.

Cette méthode génère un arbre de résolution où l'exploration se fait en profondeur d'abord et de la gauche vers la droite.

e/ Un exemple d'utilisation de PROLOG

Le petit schéma suivant représente une portion de réseau qui va être analysée en PROLOG.



Pour déclarer que t1, t2, t3 ... sont des tuyaux, on écrira :

tuyau (t1), tuyau (t2), tuyau (t3), tuyau (t4), tuyau (t5)

La définition de la relation à deux arguments JENAMONT, permettra d'indiquer quels tuyaux sont juste en amont les uns des autres :

**JENAMONT(t1,t2)
JENAMONT(t1,t3)
JENAMONT(t2,t3)
JENAMONT(t4,t3)**

L'ensemble décrivant les tuyaux et leurs relations constitue la base de faits.

PROLOG permet d'interroger cette base de faits. A la question :

Quel est le tuyau juste en amont de t3 ?

ce qui en PROLOG s'écrit : **JENAMONT(x,t3) ?**

PROLOG répondra : **x = t2**

PROLOG permet l'exploitation de la base de faits grâce à la création de nouvelles règles définissant de nouvelles relations, comme par exemple ULAMONT indiquant si un tuyau est une longueur en amont d'un autre :

$$\text{ULAMONT (x,z)} \quad \Leftarrow \quad \text{JENAMONT(x,y)} \ \& \ \text{JENAMONT(y,z)}$$

Cette règle est définie pour les variables logiques x, y, z et PROLOG recherchera automatiquement les instanciations si on lui pose la question **ULAMONT (x,z) ?**

A cette question précise, PROLOG répondra par deux couples de valeurs :

$$(x = t1, z = t3)$$

(x = t1, z = t5)

L'ensemble des règles constitue la base de connaissances du système expert.

C/ Un générateur dans un logiciel intégré : GURU

1/ La notion de logiciel intégré

GURU combine en un seul logiciel un générateur de système-expert et les modules classiques d'un logiciel intégré, c'est-à-dire :

- une base de données relationnelle avec son langage d'interrogation (SQL),
- un tableur,
- un générateur de formes pour construire les masques de saisie ou de présentation des données,
- un générateur de graphiques (courbes, diagrammes à barres..),
- un traitement de texte.

Dans le système-expert, il est possible d'utiliser les fonctionnalités de ces modules, qui ne seront pas détaillées ici.

2/ Le développement d'un système-expert avec GURU

Une base de connaissances dans GURU est formée d'un ensemble de règles, de type 'IF/THEN', composées d'une prémisse et d'une conclusion. A ces règles peuvent être attachés des commentaires qui apparaîtront à l'écran quand les utilisateurs demanderont en cours de session, à GURU de justifier un raisonnement.

a/ La prémisse d'une règle

La prémisse est constituée par un ensemble de conditions reliées par les opérateurs booléens "ET", "OU", et "XOU" représentant le OU exclusif.

Dans chaque condition, interviennent une ou plusieurs variables qui peuvent être numériques, booléennes ou alphanumériques (chaîne de caractères) ; ces variables peuvent être spécifiques au système-expert, mais elles peuvent aussi être constituées par une cellule de tableau, une variable statistique, ou le champ d'un enregistrement de la base de données.

L'écriture d'une condition peut également faire appel à des opérateurs relationnels ($<$, $=$, $>$, ...) ou numériques (+, -, * , ...) à des fonctions numériques (sin, log...) ou de traitement de chaîne de caractères, et enfin aux opérateurs booléens "NON" et "APPARTENANCE".

Voici quelques exemples de conditions gérées par GURU.

- * Si $x = 10$
- * Si $(2x + 3) < \sin y$
- * Si CONTEXTE = "crise"
- * Si MARCHE-ELEMENT-1 VRAI (Si MARCHE-ELEMENT-1)
- * Si NON MARCHE-ELEMENT-1
- * Si FILTRE-1.age > 10

où age est un champ de l'enregistrement filtre dans une base de données.

Dans GURU, toute variable doit être instanciée au moment où la condition est évaluée.

GURU ne sait pas raisonner avec des collections d'objets, et il est impossible d'écrire des règles du type :

- * S'il existe x , tel que $\text{Filtre}(x).\text{age} > 10$
- * Si quelquesoit x , $\text{Filtre}(x).\text{age} > 10$

b/ La conclusion d'une règle

La conclusion d'une règle de GURU inclut la série de commandes suivantes dont la liste n'est pas exhaustive :

- affectation de valeur à des variables, à des tableaux ou à des champs d'enregistrements de la base de données,
- stockage d'enregistrement dans la base de données,
- exécution d'un programme,

- génération d'un graphique,
- stockage d'enregistrements dans un fichier.

GURU, qui est basé sur la logique des propositions, ne peut pas gérer les relations entre variables ni générer de nouvelles relations. Les règles suivantes sont impossibles à écrire :

* Si PLUS-GRAND(x, y) & PLUSGRAND(y,z)	=>	PLUSGRAND(x,z)
* Si PERE(x,y) & PERE(y,z)	=>	GRAND-PERE(x,z)

3/ Le contrôle dans GURU

GURU explore une base de règles en chaînage avant ou arrière.

a/ En chaînage avant

Le principe du raisonnement est la saturation de la base de faits.

GURU teste les unes après les autres les N règles constituant une base de connaissance. Si, pendant ce premier cycle, au moins une règle a été appliquée, GURU enlève de la pile des règles candidates les règles appliquées, et commence un second cycle de la base de connaissances. Ce processus est recommencé autant de fois qu'au moins une règle candidate est appliquée.

b/ En chaînage arrière

GURU examine le but établi, et, en fonction de celui-ci explore la base de règles en profondeur d'abord.

c/ La métaconnaissance

La métaconnaissance dans GURU est essentiellement constituée par la gestion de variables qui déterminent l'ordre de parcours des règles ou la méthode d'évaluation des prémisses. Ainsi, l'ordre de parcours des règles peut être :

- l'ordre d'écriture,
- l'ordre des priorités définies dans chaque règle au moment où elle est écrite,
- l'ordre minimisant le nombre de variables à évaluer,
- un ordre du au hasard.

d/ La trace du raisonnement

GURU garde la trace du raisonnement et peut la présenter avec différents niveaux de détail : c'est-à-dire soit uniquement les règles appliquées, soit toutes les règles considérées, ou encore les règles considérées et les buts provisoires.

D/ Un générateur basé sur la gestion d'objets : NEXPERT-object

1/ Les grands concepts de Nexpert-object (désigné plus loin dans le texte par NEXPERT)

a/ L'introduction d'une hypothèse dans les règles

Dans les générateurs de système-expert, la syntaxe des règles est la suivante : Si (prémisses) => (actions)

Dans Nexpert, une variable booléenne, appelée hypothèse, a été ajoutée entre la partie prémisses et la partie actions. La syntaxe est donc :

Si (prémisses) alors (HYPOTHESE : VRAI) DECLENCER (actions)

b/ La formation d'un réseau reliant les règles

Une hypothèse peut constituer la prémisse d'une autre règle, aussi est-il possible de relier les règles entre-elles pour les hypothèses, pour constituer un réseau, et ainsi de visualiser un raisonnement.

Ce réseau peut être utilisé à la fois par l'expert au moment où il développe la base de règles, mais aussi par l'utilisateur du système qui peut s'en servir pour visualiser la trace du raisonnement, étant donné que les hypothèses validées, rejetées ou inconnues peuvent être représentées par 3 couleurs ou graphismes différents.

c/ La notion d'ilot de connaissances

Dans NEXPERT, un ilot de connaissances est formé par un ensemble de règles qui sont reliés par les hypothèses ou qui utilisent les mêmes données.

Cette notion d'ilot est important pour comprendre les stratégies de contrôle de raisonnement dans Nexpert.

d/ Les objets et leurs particularités

Tout objet dans NEXPERT est caractérisé par un ensemble de propriétés, par exemple, booléennes, numériques ou multivaluées.

Pompe.marque :	"RATEAU"	(variable multivaluée)
Pompe.puissance :	500	(variable numérique)
Pompe.en-marche :	VRAI	(variable booléenne).

Les objets peuvent appartenir à des classes, ce qui autorise les héritages de propriétés et de valeurs, et la manipulation d'objets par groupes.

Des objets, dits dynamiques, peuvent être créés en cours de session, ou si ces objets existant déjà, Nexpert propose la fonctionnalité de création ou de destruction de leurs liens avec une ou plusieurs classes.

L'organisation des classes, objets et propriétés est visualisable sur le réseau d'objet, construit à la demande.

f/ La prémisse d'une règle

Elle correspond à une conjonction de conditions du type

(opérateur argument 1 argument 2).

Pour les variables booléennes, la condition porte sur le fait qu'elle est vraie ou fausse, ce qui se traduit par YES pompe-1-en marche ou NO pompe-1-en-marche

Les variables numériques sont testées par un des opérateurs traditionnels, (<,>,<=,>=,<>), par exemple :

< pompe.puissance 2000

Les variables multivaluées peuvent être comparées à une chaîne de caractères indiquée ; ou entre-elles,

IS Pompe.marque "RATEAU"

La condition peut également porter sur l'appartenance d'un objet à une classe définie, désignée entre deux traits verticaux.

MEMBER Pompel.type (Pompes-A-VITESSE-FIXE)

NEXPERT est également capable d'évaluer une condition utilisant, pour les objets d'une classe spécifiée, les quantificateurs existentiel et universel. La condition

=< <POMPES-A-VITESSE-FIXE>.puissance 5000

signifie "Existe t-il un élément de la classe POMPES-A-VITESSE-FIXE avec une puissance inférieure ou égale à 5 000.

La condition

IS {POMPES-A-VITESSE-FIXE}.marque "RATEAU"

signifie "Est-ce que toutes les pompes à vitesse fixe sont de marque "RATEAU".

NEXPERT créant une liste de tous les éléments répondant aux critères demandés, il est possible de créer une seconde liste qui sera comparée à la première. En combinant les syntaxes, on peut sélectionner une liste d'objets possédant plusieurs critères, ou un seul de ces critères.

g/ La partie action d'une règle

Quand toutes les conditions d'une règle ont été vérifiées, l'hypothèse bouléenne constituant le nom de cette règle est mise à "VRAI" et les actions définies sont déclenchées. Parmi celles-ci, on peut trouver :

- une affectation de valeur à une variable bouléenne, numérique et multivaluée,
- la création ou la destruction d'un objet dynamique,
- la lecture ou l'écriture dans un fichier de données,
- l'exécution d'un programme compilé,
- le chargement d'une autre base de connaissances,
- la modification de la stratégie courante d'exploitation de la base de connaissances.

NEXPERT utilise les propriétés des objets et leur appartenance à une classe, mais par contre, ne peut pas manipuler de relations entre objets, ni en définir de nouvelles.

Comme GURU, NEXPERT n'est donc pas un générateur d'ordre 1.

2/ Le contrôle du raisonnement dans NEXPERT

Toutes les règles constituant une base de connaissances ne sont généralement pas indépendantes et NEXPERT distingue trois types de liens entre les règles :

a/ Les liens entre les règles

Lorsque l'hypothèse d'une règle constitue une condition d'une autre règle, ces deux règles sont dites liées par un lien de premier ordre, visualisé par un trait plein sur le réseau.

Lorsque deux règles utilisent une même donnée dans leurs prémisses, ou lorsque l'une modifie une donnée qui présume de l'autre, elles sont liées par un lien de second ordre.

Un ensemble de règles, reliées par des liens de premier ou de second ordre forme un ilot de connaissance. Ces ilots peuvent être reliés entre eux par des liens faibles, dits de contexte visualisés en pointillé sur le réseau.

b/ Les stratégies de raisonnement

Le fonctionnement du générateur repose sur la constitution, à chaque étape, du raisonnement, d'un agenda d'hypothèses à évaluer.

Les hypothèses prioritaires sont celles obtenues par les liens du premier ordre, puis viennent celles qui sont envisagées suite à la propagation des données, et enfin les hypothèses mises dans l'agenda par un lien de contexte.

Différentes options portant sur les liens par hypothèses, propagation de données ou de contexte, permettent de modifier les stratégies de raisonnement, pour :

- une recherche exhaustive ou non exhaustive dans l'exploitation des règles conduisant à une même hypothèse,
- la propagation immédiate ou la non propagation des changements de valeurs des données,
- le passage systématique ou conditionnel d'un ilot de connaissance à un autre.

c/ Un raisonnement en chaînage avant ou arrière

NEXPERT raisonne en chaînage arrière en fixant comme but une hypothèse à évaluer.

NEXPERT raisonne également en chaînage avant, et l'initialisation du raisonnement s'opère en fixant des valeurs à des variables, ce qui déclenche une propagation par les données.

NEXPERT, dans l'un et l'autre cas, garde en mémoire les différentes étapes du raisonnement, lequel peut être également visualisé sur les réseaux des règles et des objets.

E/ Un système basé sur la logique des prédicats d'ordre 1 : l'EXPERT-KIT

1/ Les concepts de base

L'EXPERT-KIT, écrit en Le Lisp par J. FERBER et commercialisé par la société ACT-Informatique, est basé sur la logique des prédicats et fonctionne en chaînage avant.

Construit comme un sur-ensemble de Lisp, l'Expert-Kit offre la possibilité d'utiliser des fonctions Lisp dans les règles, donc de faire appel à la bibliothèque des fonctions référencées dans le langage Le Lisp.

2/ La représentation de la connaissance dans l'EXPERT-KIT

La connaissance est représentée par des règles de production gérant des variables.

a/ La base de faits

Les faits sont représentés sous forme de listes, par exemple :

```
(pompe e1 pompe-a-vitesse-fixe "RATEAU" 5 000)
    pour décrire la pompe e1
(réservoir R1 10 40)
    pour décrire le réservoir R1
```

b/ La base de règles

Les règles sont caractérisées par un nom, une conjonction de prémisses qui seront évaluées dans l'ordre d'écriture, et une conjonction d'actions qui seront déclenchées dans l'ordre également.

c/ Les prémisses d'une règle

Si la prémisse de la règle ne contient pas de variables, le moteur d'inférences teste la présence du fait dans la base de faits. Par exemple, la règle : "Si (suite) ALORS..." sera appliquée si le fait (suite) existe.

Si la prémisse de la règle contient une ou plusieurs variables, repérées par un point d'interrogation devant leur nom, le moteur cherche à les mettre en correspondance avec un membre de la base de faits avant d'évaluer la condition sur chaque variable en utilisant des prédicats prédéfinis.

```
Si (pompe ?nom ?type ?marque ?puis)
et (member ?type (pompe-à-vitesse-fixe))
    test d'appartenance à une classe avec la fonction Lisp member.
et (< ?puissance 8000)
    opérateur d'inégalité du langage Lisp.
```

ALORS...

Avec cette prémisse, le moteur d'inférences construit la liste des pompes à vitesse fixe, de puissance inférieure à 8000, présentes dans la base de faits, et appliquera la règle pour chaque élément de cette liste.

Si fonction Le-Lisp utilisé dans une règle n'appartient pas à la bibliothèque standard, elle doit être décrite dans un fichier qui sera chargé et évalué en même temps que la base de connaissances.

d/ Les actions d'une règle

Les actions peuvent avoir comme conséquences, outre les fonctions de lecture et d'affichage :

- ajout, suppression ou la modification d'un fait,
- application d'une fonction LISP, de la bibliothèque standard, ou décrite dans un fichier,
- manipulation de listes d'éléments,
- modification des stratégies de contrôle par activation ou inhibition de paquets de règles.

3/ Les stratégies de contrôle dans l'EXPERT-KIT

L'Expert-Kit fonctionne en chaînage avant par saturation de la base de faits, en examinant les règles dans l'ordre où elles ont été écrites.

Une possibilité d'agir sur le raisonnement est la structuration des règles en paquets, et l'activation et l'inhibition de ces paquets à l'aide des prédicats précédemment évoqués : "ACTIVER" et "DESACTIVER".

F/ Un générateur utilisant la logique des prédicats d'ordre 2 : SNARK-OPEN

1/ Les concepts de base

Le fonctionnement de ce générateur est basé sur la logique des prédicats d'ordre 2, ce qui lui permet de raisonner sur les propriétés des relations, et de créer et manipuler des classes d'entités.

SNARK a comme fonctionnalité la gestion d'hypothèses, ou "backtrack", pour explorer les solutions possibles d'un problème. L'hypothèse, dans ce cas, n'est pas une variable booléenne identifiant une règle, comme dans NEXPERT mais une instantiation ou une affectation d'un fait dans la base de faits.

SNARK a aussi des stratégies de contrôle caractéristiques, comme l'activation ou la désactivation de paquets de règles, et l'utilisation de démons établissant des priorités absolues pour certaines règles.

La représentation des connaissances dans SNARK est formalisé sous forme de règles de production et les éléments de connaissance dans les règles et les faits sont des triplets. La description d'un problème est donc une liste de relations binaires.

2/ La connaissance dans SNARK-OPEN

a/ La base de faits

Les faits sont représentés dans SNARK-OPEN par un triplet <entité, entité, entité> dans lequel la première entité désigne un objet ou un individu du domaine étudié, la deuxième entité désigne un attribut, et la troisième entité désigne une valeur.

Un triplet s'écrit donc sous la forme :

< objet, attribut, valeur >, mais la caractéristique de SNARK est qu'une même entité peut être un objet dans un fait, une valeur dans un autre fait, et un attribut dans un troisième fait. Par exemple, l'entité "**plus puissante que**" est :

- un attribut	dans le fait	pompe E1 plus-puissante pompe-e11.
- un objet	dans le fait	plus-puissant propriété transitive.

L'entité pompe-e11 est une valeur dans le premier fait, et un objet dans le fait : **pompe-e11** **marque** **"RATEAU"**

De plus, chaque fait peut être affecté d'un coefficient de vraisemblance compris entre 0 et 1.

b/ La base de règles

La base de règles est formée d'un ensemble de règles repérées par des noms, et constituées d'une conjonction de prémisses suivie d'une conjonction d'actions, qui seront exécutées dans l'ordre, lors de l'application de la règle.

La syntaxe classique d'une règle est donc :

nom-de-la-règle :

Si (prémisses)
ALORS (actions)

Il existe un autre type de règle dont la syntaxe est :

DES QUE (prémisses)
ALORS (actions)

Dans les règles, il est possible d'utiliser des variables qui seront mises en correspondance par le moteur d'inférences avec des entités de la base de faits.

L'emploi d'une variable uniquement dans les prémisses d'une règle est la traduction dans SNARK du quantificateur existentiel (il existe).

c/ Les prémisses d'une règle

Une règle devient applicable dès que toutes ses prémisses peuvent être mises en correspondance avec des faits présents dans la base de faits.

SNARK OPEN distingue les prémisses positives, négatives et conditionnelles.

Pour les prémisses positives, la syntaxe est la suivante :

(<objet> <argument>) (opérateur) (objet N(<objet> <argument>))

Voici quelques exemples :

Si TYPE (POMPE) = pompe-à-vitesse-fixe
MARQUE (POMPE) = "RATEAU"
PUISSANCE (POMPE-AVAL) > PUISSANCE(POMPE-AMONT)

Les prémisses négatives peuvent être utilisées pour exprimer des conditions portant sur l'absence de faits dans la base, par exemple :

Si NON (fonctionnement (Pompe) = arrêt)

Les prémisses conditionnelles expriment des conditions sur des variables déjà utilisées dans d'autres prémisses, ce qui fait qu'elles auront déjà été instanciées, par exemple :

$$(X) > 2000$$

$$(X) \leq (Y)$$

d/ Les actions d'une règle

Outre les opérations de lecture, écriture et affichage, les actions exécutées lors de l'application d'une règle sont de différente nature.

SNARK peut affecter une valeur selon 2 opérations qui ont des effets différents. La première opération va ajouter un fait dans la base de faits, ce qui peut conduire à une variable multivaluée tandis que dans le second type d'opération d'affectation, toute autre valeur de la variable, est détruite, et cette variable ne peut être que monovaluée.

SNARK offre la possibilité de créer ou supprimer des entités dans la base de faits, ou d'effacer tous les faits correspondant à un couple objet-attribut déterminé .

Enfin, l'action d'une règle peut consister à reprendre le raisonnement à un état antérieur, en remettant la base de connaissances dans l'état où elle était avant (examen de l'hypothèse rejetée).

Notons que le déclenchement d'une règle peut arrêter immédiatement le raisonnement, ou désactiver définitivement une règle.

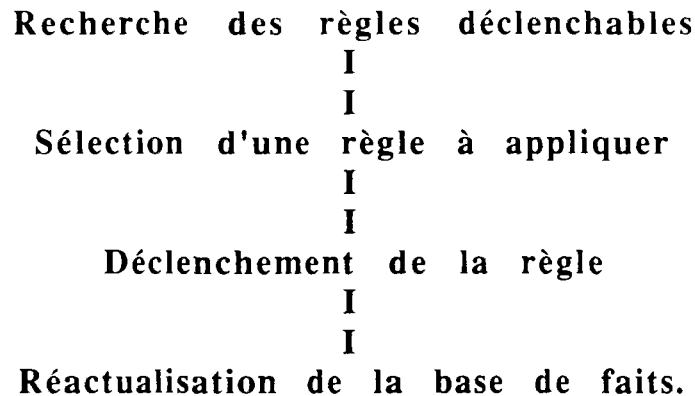
3/ Les stratégies d'exploration de la base de règles le contrôle :

a/ Un raisonnement en chaînage avant par saturation

Au début d'une session, les seules règles déclenchables sont celles dont les prémisses peuvent être mises en correspondance avec la base de faits initiale.

Parmi ces règles, une seule sera appliquée, et son exécution réactualisera la base de faits. Cette règle est ensuite enlevée de l'agenda des règles déclenchables, jusqu'à ce que des faits, intervenant dans ses prémisses, soient modifiés.

L'exploration de la base de règles reproduit ce cycle qui se schématise ainsi :



b/ Le contrôle dans la sélection des règles

Le processus de sélection des règles peut être modifié par création de paquets de règles, par les jeux d'actions consistant à activer ou inhiber certaines règles, et par l'utilisation des règles démons.

c/ La gestion d'hypothèse

Une action particulière permet à SNARK de ramener la base de connaissance dans un état antérieur du raisonnement, en utilisant la trace du raisonnement.

Ainsi, tous les faits créés ou détruits, depuis un point spécifié dans la règle, sont supprimés, et le moteur peut alors examiner, à partir de ce point, une autre hypothèse.

d/ La logique d'ordre 2

La logique d'ordre 2 apporte la possibilité de raisonner avec les propriétés des relations. A partir de la base de faits :

pompe-1	en-amont	pompe-2
pompe-2	en-amont	pompe-3
en-amont	propriété	transitive

et de la règle :

Si propriété (X) = transitive

(X) (a) = (b)

(X) (b) = (c)

ALORS

(X) (a) = (c) (les arguments de la propriété X sont a et b).

On pourra déduire : **pompe-1** **en-amont** **pompe-3,**

La logique d'ordre 2 permet aussi de manipuler des classes d'entités qui peuvent être non seulement des objets, mais aussi des attributs. Les classes d'entité servent à gérer l'héritage ou les vérifications de cohérence entre les objets.

pompe	élément-description	marque
pompe	élément-description	puissance
réservoir	élément-description	hauteur
réservoir	élément-description	surface
pompe e-1	marque "RATEAU"	
réservoir-R1	hauteur 4 m	

Avec la base de faits précédente, la logique d'ordre 2 permet de définir des propriétés communes aux classes d'objets "pompe" et "réservoir", mais aussi à l'attribut "élément-description".

G/ Quelques éléments sur les générateurs pour mini-ordinateurs

Les générateurs comme KEE, ART, PICON, fonctionnant sur mini-ordinateurs ont comme caractéristique essentielle l'intégration des fonctions de représentation graphique au système.

Cela signifie que la visualisation des variables sera faite en temps réel, et tout changement de valeur pour une variable donnée, sera donc immédiatement prise en compte dans sa représentation graphique.

ANNEXE 2

Δ Δ Δ

**Programmes d'apprentissage
symbolique automatique :**

**Finalités et descriptions de
quelques algorithmes**

O O O

SOMMAIRE DE L'ANNEXE

I) APPRENTISSAGE PAR REORGANISATION DE LA CONNAISSANCE	page 3
A) L'idée générale	
B) L'apprentissage par implantation	
C) L'apprentissage par reconfiguration progressive d'un modèle	
II) APPRENTISSAGE PAR INSTRUCTION	page 7
A) L'idée générale	
B) Les phases d'une session d'apprentissage	
C) Exemple	
III) APPRENTISSAGE DEDUCTIF PAR RECHERCHE D'EXPLICATION	page 8
A) L'idée de base	
B) Les étapes d'une session d'apprentissage	
C) Exemple	
D) Intérêts et limites de ce type d'apprentissage	
IV) APPRENTISSAGE INDUCTIF A PARTIR D'EXEMPLES	page 10
A) L'idée générale	
B) L'espace des instances	
C) L'exploration de l'espace des règles (hypothèses)	
D) L'apprentissage d'un concept unique	
V) APPRENTISSAGE PAR EXPERIMENTATION	page 24
A) L'idée générale	
B) Les étapes d'une session d'apprentissage	
C) Le programme d'intégration symbolique LEX	
D) La génération et le test d'hypothèses : le programme DENDRAL	

VI) APPRENTISSAGE INDUCTIF A PARTIR D'OBSERVATIONS

page 30

- A) L'idée générale
- B) La découverte de lois : le programme BACON
- C) La recherche de similitudes

VII) APPRENTISSAGE PAR ANALOGIE

page 41

- A) La signification du terme
- B) L'analogie relationnelle
- C) La formation de concepts
- D) L'analogie transformationnelle
- E) L'analogie dérivationnelle

I) APPRENTISSAGE PAR REORGANISATION DE LA CONNAISSANCE

A) L'IDEE GENERALE :

Cette forme d'apprentissage s'apparente à de la mémorisation, c'est-à-dire à l'enregistrement de connaissances nouvelles de telle sorte que, en cas de besoin, elles puissent être facilement retrouvées et évitent des calculs, une série d'inférences ou des requêtes effectués précédemment.

Le schéma général est le suivant. Une fois qu'un problème particulier a été résolu par le système, le module d'apprentissage enregistre le problème et sa solution, sous forme de paires, avec un système d'indexage.

B) APPRENTISSAGE PAR IMPLANTATION :

1) Description de ce programme :

Un système utilise une fonction f qui associe à un vecteur entrée $(X_1 \dots X_p)$ un vecteur sortie $(Y_1 \dots Y_p)$. L'implantation de connaissance consiste alors à stocker les paires associées $[(X_1 \dots X_p), (Y_1 \dots Y_p)]$ avec un indexage basé sur $(X_1 \dots X_p)$. Toute session ultérieure du système commence par une exploration de la mémoire en vue d'économiser un calcul.

Le programme de jeu d'échecs, développé par SAMUEL, utilise cette méthode pour éviter de recalculer des valeurs attribuées à des configurations de jeu.

2) Intérêt et limites de ce type d'apprentissage :

Même si le principe de cet apprentissage paraît très simple, l'archivage des paires nécessite une organisation stricte pour que le gain obtenu en diminuant les calculs ne soit pas perdu lors de l'exploration de la mémoire. L'intérêt de l'implantation étant avant tout la vitesse, les temps d'exploration des situations stockées doivent donc être nettement inférieurs à ceux des calculs qu'ils sont censés éviter.

Pour cela, les programmes possèdent des modules de hiérarchisation des index, pour retrouver rapidement un problème déjà traité identique à celui considéré. Ces programmes opèrent également des regroupements de problèmes similaires, par exemple en recherchant les symétries, pour diminuer le nombre d'objets archivés distincts.

C) APPRENTISSAGE PAR RECONFIGURATION PROGRESSIVE D'UN MODELE :

1) L'idée de base :

Ce type de programme d'apprentissage, développé à l'université VRIJ de Bruxelles par M Van de VELDE, s'applique aux systèmes experts dans lesquels deux niveaux de connaissance ont été définis : une connaissance profonde et une connaissance superficielle ou usuelle.

a) La connaissance profonde :

Cette connaissance est basée, soit sur un modèle structurel décrivant l'ensemble des composants et leurs relations, soit sur un modèle fonctionnel simulant le comportement de l'appareil à partir de celui de ses composants, soit encore sur un modèle causal connaissant l'ensemble des relations entre les propriétés de ses composants.

Ce type de modèle doit être le plus complet possible. Il en résulte qu'il pourra décrire, simuler ou expliquer un comportement dans pratiquement tous les cas, mais cette exhaustivité a le plus souvent comme contre-partie un manque de rapidité dans les réponses.

b) La connaissance usuelle :

Cette connaissance est essentiellement basée sur l'expérience et n'est pas strictement formalisée. Les heuristiques sur lesquelles elle s'appuie permettent, dans la plupart des cas, d'accéder rapidement à une réponse ou une décision.

c) La finalité de l'apprentissage :

Les algorithmes d'apprentissage cherchent à relier, en s'appuyant sur l'expérience accumulée par le système informatique, les deux types de connaissances. Il s'agit en quelque sorte de faire émerger la partie de la connaissance profonde qui est effectivement utilisée, et d'ordonner l'espace de recherche de façon à faire apparaître rapidement la solution dans tous les cas usuels.

2) Méthodologie de l'apprentissage à partir d'un modèle causal :

a) Structuration de la connaissance profonde :

La méthodologie va être expliquée en prenant exemple sur un arbre ou un graphe expliquant un symptôme observé par un diagnostic, illustré ci-dessous.

SYMPTOME			

-----	-----	-----	-----
vérification a	vérification b	vérification c	vérification d
-----	-----	-----	-----
-----	-----	-----	-----
diagnostic a	diagnostic b	diagnostic c	diagnostic d
-----	-----	-----	-----

Le modèle causal opère une recherche systématique et aveugle dans l'espace de recherche et il apparaît vite que toutes les questions posées et vérifications demandées par le système ne sont pas pertinentes et qu'il serait très utile de les minimiser. C'est ce que fait l'algorithme d'apprentissage à partir des cas réellement rencontrés par le système.

b) La recherche des règles opérantes sur le cas courant :

Cette phase, consécutive à la validation d'un diagnostic, consiste à rechercher les règles qui mettent en relation le plus directement possible le symptôme avec le diagnostic validé.

c) L'intégration des règles opérantes :

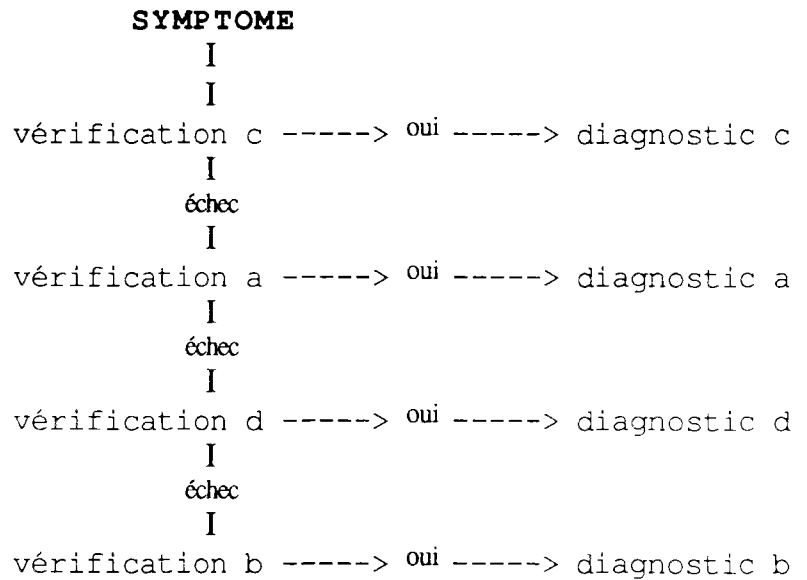
La finalité de cette phase est la résolution des conflits entre règles opérantes, pour ordonner l'ordre d'application des règles.

Cet ordre peut être défini à partir de statistiques sur la fréquence d'utilisation des règles, ou de coût d'utilisation de règles (certaines vérifications sont beaucoup plus onéreuses que d'autres), ou encore selon d'autres critères à définir.

d) La modification du modèle causal :

L'intégration ne rend invalide aucune règle, mais restreint seulement leur applicabilité.

Il en résulte que l'organisation du modèle causal est alors modifiée par l'expérience. Par exemple, après l'étude d'une centaine de cas, le modèle causal décrit précédemment pourrait devenir.



3) En conclusion :

Les deux phases de sélection et d'intégration des règles concourent pour construire une mémoire compacte de l'expérience passée. Certaines méthodes de généralisation décrites précédemment peuvent s'appliquer ici, et augmenter le pouvoir des règles.

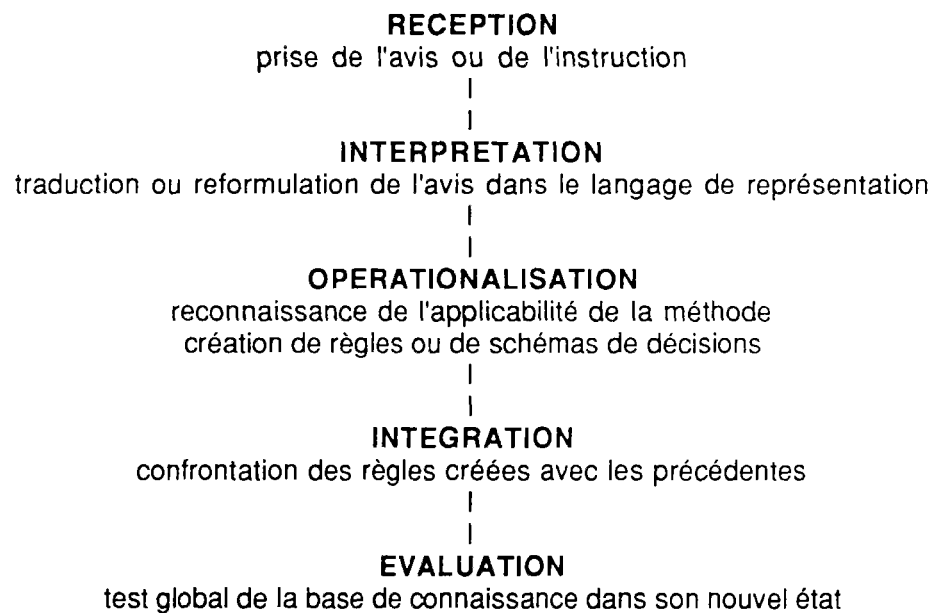
II) APPRENTISSAGE PAR INSTRUCTION

A) L'IDEE GENERALE :

L'approche présentée dans ce chapitre décrit des systèmes qui intègrent des avis ou instructions venant d'un "maitre" extérieur au système et les convertissent en règles qui seront applicables par le module de réalisation du système.

B) LES PHASES D'UNE SESSION D'APPRENTISSAGE :

Une session complète se déroule en cinq étapes, selon le schéma ci dessous :



1) La réception

Cette phase, généralement passive, est active pour certains systèmes qui définissent la partie de la base de connaissance sur laquelle ils souhaitent recevoir des avis.

2) L'interprétation :

Cette étape consiste à transformer sans déformation et sans perte l'ensemble des informations contenues dans l'instruction dans le langage de représentation interne du système.

3) L'opérationnalisation :

Une instruction, donnée à un niveau de généralité élevée est rarement utilisable en tant que telle par le système. Elle doit être adaptée aux différents contextes et cas particuliers que le système traite couramment.

L'opérationnalisation fait appel à des méthodes internes au système dont la fonction est de développer des hypothèses sur la signification de l'instruction pour l'ensemble des cas que le système doit apprendre à résoudre.

Ces hypothèses portent, selon le langage de représentation du système, sur la création ou la modification de règles ou de schémas de raisonnement. Une fois définies, ces hypothèses sont testées et corrigées avant d'être acceptées.

4) L'intégration :

Les modifications acceptées précédemment ne sont introduites dans le système qu'après une confrontation des règles ou schémas nouvellement créés avec les anciens.

Ce qui est nouveau ne doit en effet être ni redondant, ni en contradiction avec la base de connaissance en son état. Ainsi le recouvrement des prémisses d'une règle nouvelle avec celles d'une ancienne conduira à entamer une procédure de spécialisation ou de définition de métarègles. De même une contradiction provoquera une modification des actions ou à la formulation d'une requête au "maitre" dans l'attente d'une nouvelle instruction levant la contradiction.

5) L'évaluation :

La dernière phase consiste à tester l'amélioration des performances de la base de connaissance après la phase d'intégration.

III) APPRENTISSAGE DEDUCTIF PAR RECHERCHE D'EXPLICATION

A) L'IDEE DE BASE :

Le terme recherche d'explication est employé parce que la méthodologie suivie par le programme est la découverte des étapes expliquant pourquoi un objet particulier, présenté par un "maitre", satisfait un concept donné. Une fois l'explication trouvée, le programme cherche à décrire son champ de validité le plus large, remplaçant ainsi une définition fonctionnelle d'un concept, nécessitant à chaque fois d'en faire la preuve, par une définition structurelle, décrivant ses conditions de validité.

La recherche d'explication s'opère dans la théorie du domaine, définie comme l'ensemble des axiomes et règles d'inférences utilisables pour expliquer comment une instance particulière illustre un concept ou une relation déterminée.

Ce type d'apprentissage porte aussi le nom d'apprentissage analytique, parce qu'il fait appel à l'analyse de la trace du raisonnement établissant la preuve du concept .

B) LES ETAPES D'UNE SESSION D'APPRENTISSAGE :

L'approche traditionnelle comprend deux phases. La première est la construction de l'explication prouvant que l'exemple donné est bien une instance positive du concept but, à partir des axiomes

et règles de la théorie du domaine. La trace de l'explication se présente sous la forme d'un arbre dont les branches convergent vers le concept but.

La seconde étape utilise cette trace pour définir l'ensemble des conditions suffisantes portant sur les descriptions des objets, pour lesquelles l'explication sera toujours valide. La recherche de la version la plus générale des descriptions est réalisée par une analyse de l'arbre de la trace, en remontant du concept but vers les branches en remplaçant les constantes par des variables et en unifiant les variables chaque fois que la théorie du domaine le permet.

C) EXEMPLE D'UNE SESSION D'APPRENTISSAGE :

1) La théorie du domaine :

```
transfert_de_vers(X,Y) <=  superieur(pression(X), pression(Y))&reliés(X,Y)
decroit(vol(X)&croit(vol(Y)) <=  transfert_de_vers(X,Y)
baisse_niveau(X) <=  decroit(vol(X))
monte_niveau(X) <=  croit(vol(X))
```

2) Description de l'instance positive :

```
reliés(A,B)
superieur(pression(A), pression(B))
baisse_niveau(A)
```

3) Construction de l'arbre de la preuve (trace du raisonnement)

```
superieur(pression(A), pression(B))
                                &                => transfert_de_vers(A,B)
reliés(A,B)

transfert_de_vers(A,B)  =>  decroit(vol(A)&croit(vol(B))
decroit(vol(A) => baisse_niveau(A)
```

4) La généralisation du raisonnement :

Ce raisonnement se généralise en remplaçant les instances A par la variable X et B par la variable Y.

5) L'ensemble des conditions suffisantes du concept "baisse-niveau" :

```
reliés(X,Y)
superieur(pression(X), pression(Y))
```

D) INTERET ET LIMITE DE CE TYPE D'APPRENTISSAGE :

Cette approche analytique est purement déductive et n'ajoute pas d'informations. Son rôle est de transformer des définitions fonctionnelles implicites de concepts, nécessitant à chaque utilisation un raisonnement, en définitions structurelles explicites et opérationnelles pour les sessions ultérieures du système.

IV) APPRENTISSAGE INDUCTIF A PARTIR D'EXEMPLES :

A) L'IDEE GENERALE :

1) Définition et finalité de l'apprentissage inductif :

L'apprentissage inductif est le procédé d'acquisition de connaissances à partir d'inférences inductives sur des faits fournis par l'environnement.

Les programmes qui apprennent à partir d'exemples doivent extraire d'une collection d'instances spécifiques des règles générales utilisables ensuite par le module de réalisation pour classer des objets selon un ou plusieurs concepts. Le terme concept est alors défini comme un prédicat, exprimé dans un langage de description donné, considéré comme vrai pour chaque exemple et comme faux pour chaque contre-exemple.

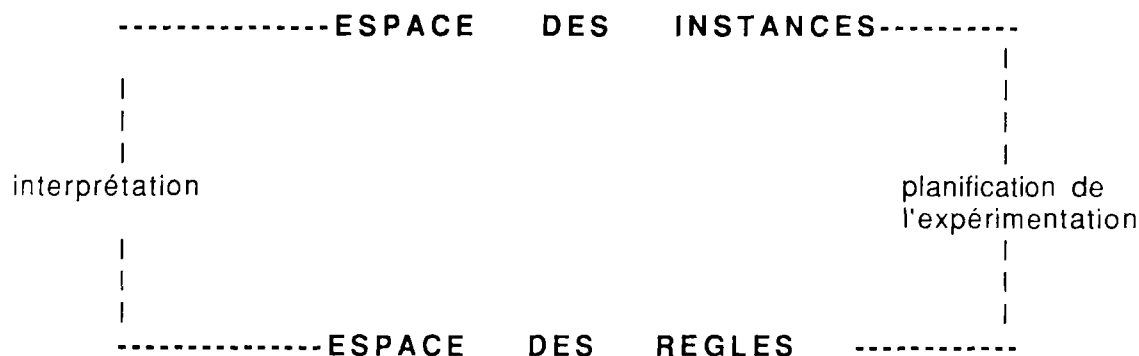
La tâche d'apprentissage est donc l'acquisition d'un ou plusieurs concepts, préalablement définis, dans un ensemble d'instances formées d'exemples et, dans certains cas, de contre-exemples.

Ces programmes sont différents de ceux qui apprennent à partir d'observations, dont la finalité est la structuration d'un ensemble d'instances, pour lesquelles aucun concept n'a été défini préalablement. Cette finalité se rapproche de celle de certaines techniques de l'analyse des données, comme les classifications par exemple.

Ces instances spécifiques, présentées au programme, représentent un niveau très bas d'informations à généraliser avant d'être exploitées efficacement, pour déterminer si une instance présentée au programme après la session d'apprentissage appartient ou non au concept.

2) Dualité des espaces de règles et d'instances :

Dans la plupart des cas, le problème d'apprentissage peut être vu comme une exploration combinée de 2 espaces :



Les informations fournies au système sont contenues dans l'espace des instances sous forme d'ensembles d'exemples et contre-exemples, caractérisés par un ensemble de descripteurs.

L'interprétation commence par une analyse des descripteurs d'une ou plusieurs instances pour extraire l'information utile à une exploration pertinente de l'espace des règles correspondant à la tâche à réaliser.

B) L'ESPACE DES INSTANCES :

1) Les structurations possibles :

L'espace des instances est structuré selon la nature de l'information qui est fournie au système, et la tâche d'apprentissage à effectuer. Les programmes travaillent dans quatre types d'espaces différents qui peuvent constituer un ensemble formé :

- d'une collection d'exemples d'un seul concept,
- d'une collection d'exemples et contre-exemples d'un seul concept,
- de plusieurs groupes disjoints d'exemples de différents concepts,
- de plusieurs groupes d'exemples de concepts pouvant se superposer.

2) La caractérisation des instances : la notion de descripteur :

2.a) Les différents types des descripteurs :

On note trois types de base parmi les descripteurs :

Les premiers sont les descripteurs nominaux ou de catégories, représentés par des symboles ou des noms sans aucune structure entre les différentes valeurs, comme par exemple la couleur possible d'un objet.

Viennent ensuite les descripteurs linéaires, définis par une valeur numérique, comme par exemple une longueur ou une surface.

Enfin les descripteurs structurés sont généralement représentés sous la forme d'un arbre. La forme d'un objet est un exemple de ce type en ce sens qu'un carré est un rectangle, lui-même un parallélogramme.

2.b) Restriction de l'espace des descriptions par des contraintes :

Un certain nombre de contraintes, comme par exemple leur interdépendance, restreignent l'espace des descripteurs, c'est à dire l'ensemble des valeurs possibles. Dans beaucoup de problèmes, certaines variables spécifient un état et d'autres le décrivent, et il est évident que la description d'un état n'a de sens que si un descripteur indique que l'objet est dans cet état.

Les descripteurs peuvent avoir des propriétés comme la réflexivité, la transitivité ou la symétrie qui limitent le champ des valeurs possibles.

Des interrelations parmi les descripteurs imposent également des restrictions, dues aux définitions mêmes des descripteurs . Par exemple :

$$\text{longueur (objet)} > \text{largeur (objet)}$$

Dans certains cas de nouveaux descripteurs peuvent être définis en cas de besoin:

$$\text{volume (objet)} = \text{surface (objet)} * \text{hauteur (objet)}.$$

2.c) La pertinence des descripteurs initiaux :

Les descripteurs utilisés dans la plupart des cas sont des caractéristiques observables ou des mesures disponibles sur les objets. Dans une recherche de définition de concepts, tous n'ont pas le même degré de pertinence et 3 cas peuvent être distingués.

La pertinence peut être complète et tous les descripteurs directement pertinents pour la tâche d'apprentissage. La pertinence peut être partielle et les exemples contenir des descripteurs inutiles ou redondants. La tâche d'apprentissage comprend alors la sélection des plus pertinents.

Enfin la pertinence est parfois indirecte et les exemples ne contiennent pas les descripteurs directement pertinents. Cependant certains des descripteurs peuvent être utilisés pour construire des descripteurs dérivés qui seront directement pertinents. Par exemple, les objets peuvent être caractérisés par leurs dimensions alors que le descripteur pertinent est le volume.

2.d) L'annotation des descripteurs :

L'annotation d'un descripteur est une collection d'informations de base sur ce descripteur, qui peut comprendre :

- une spécification de son domaine et de son type,
- une spécification des opérateurs qui lui sont applicables,
- une spécification des contraintes et relations entre lui et les autres,
- pour les descripteurs numériques : la moyenne, la variance,
- une caractérisation des objets auxquels il est applicable,
- une spécification de la classe des descripteurs le contenant le descripteur.

3) La sélection des instances :

La planification de l'expérimentation recouvre à la fois la sélection active d'instances, ou le choix de descripteurs particuliers de ces instances.

Dans la plupart des cas, le programme ne sélectionne pas les instances mais les analyse dans l'ordre où elles lui sont fournies par l'environnement extérieur. Cependant, plusieurs programmes disposent d'heuristiques pour explorer cet espace selon plusieurs approches.

Une première approche consiste à discriminer autant que possible l'ensemble des hypothèses alternatives H , à chaque étape du problème, en sélectionnant l'instance qui invalide le plus grand nombre d'hypothèses.

Une autre approche sélectionne les contre-exemples des hypothèses de l'ensemble H , de telle sorte que le programme d'apprentissage se focalise sur les conditions d'invalidation des hypothèses.

Les heuristiques ne portent pas uniquement sur la sélection des instances, mais aussi sur l'analyse des descripteurs de ces instances.

C) L'EXPLORATION DE L'ESPACE DES REGLES :

1) La formalisation d'un concept par une description :

La finalité de l'apprentissage consiste à définir la description la plus plausible du concept dont les exemples sont donnés.

2) La notion de description :

Une description est l'ensemble des propriétés communes à une collection d'instances présentées comme autant d'exemples d'un concept par l'environnement. Une description est donc caractérisée par une conjonction de valeurs de descripteurs, ex :

ROUGE(X)&SURFACE(X,12)&RECTANGLE(X)

Parfois une description est composée d'une disjonction de plusieurs conjonctions, par exemple le concept GRAND-MERE(X,Y) est défini comme :

MERE(X,Z)&MERE(Z,Y) ou MERE(X,Z)&PERE(Z,Y).

2.a) Les descriptions caractéristiques :

Une description caractéristique d'une classe d'objets spécifie toutes les propriétés communes des objets connus dans la classe, et définit la classe dans le contexte d'un nombre non limité d'autres classes d'objets.

Par exemple, dans un ensemble d'animaux, une description caractéristique opposera la description du concept "chien" à celle de tous les autres animaux.

Parmi l'ensemble des descriptions, la description caractéristique maximale est définie comme étant la plus spécifique, c'est-à-dire la plus longue, tout en caractérisant l'ensemble des objets dans une classe donnée.

Une telle description représente, pour le concept "chien", l'ensemble maximal des propriétés communes à tous les chiens donnés en exemples. Cette description discrimine au mieux la classe donnée de toutes les autres classes possibles.

2.b) Les descriptions discriminantes :

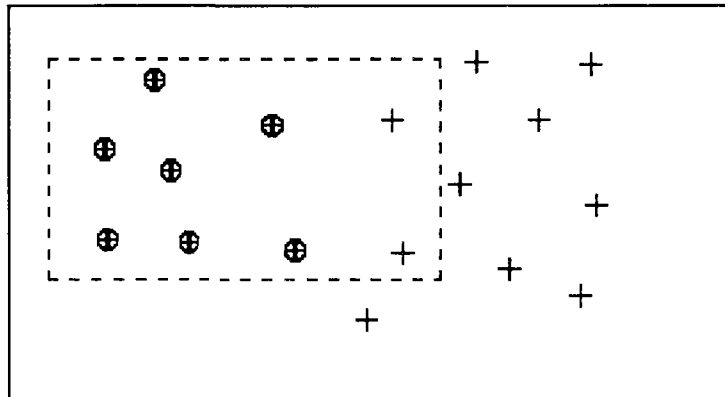
Une description discriminante d'une classe d'objets distingue la classe donnée d'un nombre limité d'autres classes.

Dans notre collection d'animaux, une telle description distinguera le concept "chien" des autres animaux appartenant à une classe définie, comme "chat", "mouton",...

Une telle description discriminante devient minimale lorsqu'elle contient le nombre minimal de descripteurs distinguant tous les objets dans une classe donnée des objets des autres classes. De telles descriptions ont pour but de spécifier l'information minimale suffisante pour identifier la classe donnée parmi un nombre fixé d'autres classes. C'est par exemple l'ensemble minimal des propriétés caractéristiques du concept "chien".

2.c) Conditions de complétude et de consistance :

Une description est dite complète si chaque élément de l'ensemble d'apprentissage de n'importe quelle classe satisfait la description de cette classe, même si des éléments couverts par la description n'appartiennent pas à la classe.

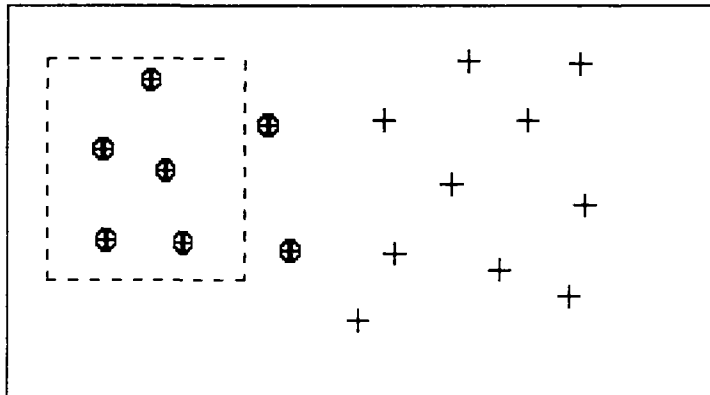


⊕ Exemple

+ Contre-exemple

⌚ Description complète

Une description est consistante quand tout élément satisfaisant la description d'une classe quelconque n'est pas membre d'une autre classe. Aucun contre-exemple n'est couvert par la description, mais celle-ci ne couvre pas obligatoirement tous les objets de la classe.



⊕ Exemple

+ Contre-exemple

⌚ Description consistante

3) Structuration de l'espace des descriptions :

3.a) Généralité et spécificité d'une description :

Une description D1 est plus générale qu'une description D2, si et seulement si, toutes les instances couvertes par la description D2 sont également couvertes par D1.

3.b) L'ordonnancement partiel de l'espace des descriptions :

La relation de généralité relative aux descriptions impose un ordonnancement partiel de l'espace des hypothèses. La structure définie est importante car elle fournit une base pertinente pour organiser l'exploration de l'espace, à partir d'une partition de l'espace par les ensembles S et G.

3.c) Les ensembles S et G :

L'ensemble S est constitué, à chaque étape de l'exploration, de toutes les hypothèses consistantes avec les instances traitées et pour lesquelles il n'existe aucune description à la fois plus spécifique et consistante.

L'ensemble G est constitué, à chaque étape de l'exploration, de toutes les hypothèses consistantes avec les instances traitées et pour lesquelles il n'existe aucune description à la fois plus générale et consistante avec les contre-exemples déjà observés.

4) Les opérateurs ou règles de généralisation :

4.a) La fonction de ces opérateurs :

La détermination des hypothèses candidates à une étape du problème peut se comprendre comme un processus d'application de règles de généralisation aux descriptions initiales et intermédiaires.

Rechercher des hypothèses candidates à partir d'instances se caractérise par une exploration heuristique d'un espace d'états où les états sont les instances avec leurs descripteurs, les opérateurs des règles de généralisation, spécialisation et reformulation et le but est la description du concept.

4.b) Méthodes de généralisation :

Ces méthodes sont décrites par 4 règles principales de généralisation :

- La règle d'altération d'une conjonction :

La description ROUGE(X)&GRAND(X)&RECTANGLE(X) produit 2 généralisations G1 et G2

G1 : ROUGE(X)&RECTANGLE(X) et G2 : GRAND(X)&RECTANGLE(X)

Cette règle établit qu'une description peut être généralisée en éliminant un des termes d'une conjonction.

- La règle d'extension d'une constante à une variable :

HOMME(paul)&GRAND(paul) se généralise en HOMME(X)&GRAND(X).

- La règle d'extension à un intervalle pour un descripteur linéaire:

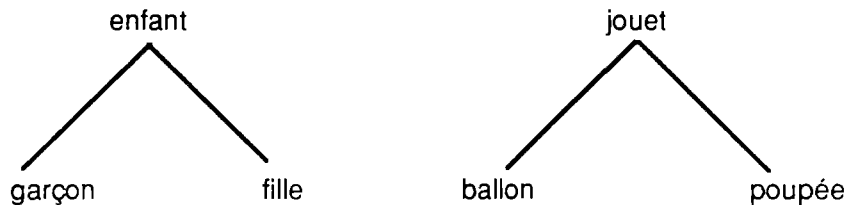
EGAL(10,A)&APPARTIENT(u,A); et EGAL(12,B)&APPARTIENT(u,B)

se généralise en : Pour tout x, tel que $x \in (10..12)$, APPARTIENT(u,X).

- La règle de remontée dans la hiérarchie conceptuelle pour un descripteur structuré :

L'exemple suivant, utilisant des descripteurs structurés, illustre cette règle :

BALLON(b1)&GARCON(paul)&JOUE-AVEC(b1,paul)
POUPEE(p1)&FILLE(julie)&JOUE-AVEC(p1,julie)



Les deux scènes sont identiques à un niveau conceptuel supérieur, c'est-à-dire :

JOUET(X)&ENFANT(Y)&JOUE-AVEC(Y,X)

Les quatre règles décrites ci-dessus utilisent les descripteurs initiaux sans en générer de nouveaux. Certains programmes possèdent des bibliothèques de règles de construction inductives qui sont capables de produire de nouveaux descripteurs et en conséquence de nouvelles généralisations. Les règles les plus simples de ce type portent sur les nombres d'éléments dans une classe et leur ordonnancement.

4.c) Règles de généralisation constructives :

Ce type de règles génère des hypothèses candidates en utilisant des descripteurs qui ne sont pas dans les instances d'origine, comme par exemple le nombre de parties constituant un objet .

5) Méthodes d'exploration de l'espace des hypothèses :

D'une manière générale, les programmes gèrent par étapes l'ensemble H des hypothèses les plus plausibles dans l'état de connaissance du système. Les méthodes diffèrent essentiellement sur les modes de gestion et sont en première analyse séparables en deux classes :

5.a) Les méthodes conduites par les données :

Les méthodes d'exploration conduites par les données prennent en compte à chaque étape les écarts entre l'hypothèse courante et les données disponibles pour déterminer les révisions à effectuer sur cette hypothèse.

Dans le cas où l'exploration a lieu en profondeur d'abord, le programme ne travaille qu'avec une seule hypothèse, celle qui est considérée comme la meilleure et qui est affinée étape par étape.

Si l'exploration se fait par largeur d'abord, le programme maintient un ensemble d'hypothèses alternatives qui sont successivement évaluées et modifiées.

5.b) Les méthodes conduites par un modèle :

Ces méthodes consistent dans la plupart des cas à générer des hypothèses à chaque étape du raisonnement et à tester ces hypothèses nouvellement générées sur l'ensemble des instances disponibles.

5.c) Un sens dans l'exploration de l'espace des hypothèses :

L'espace des descriptions peut être parcouru en prenant en compte d'abord les descriptions des instances initiales et en cherchant à les généraliser pour qu'elles restent consistantes avec les nouvelles instances.

D'autres programmes travaillent en partant de l'hypothèse la plus générale, c'est-à-dire la description nulle, et produisent des hypothèses de plus en plus spécifiques à chaque itération.

L'approche est parfois combinée lorsque le programme maintient deux ensembles d'hypothèses en travaillant par généralisation dans le premier et par spécialisation dans le second en cherchant à faire coïncider les descriptions des deux ensembles.

6) L'utilisation d'heuristiques pour guider l'exploration :

6.a) La connaissance de base du problème :

Les questions traditionnelles qui se posent lors de la définition des hypothèses candidates sont :

- quel type d'opérateurs le programme doit-il utiliser pour les générer ?
- quels sont les meilleurs critères pour guider la sélection pour évaluation ?
- comment peuvent-elles être confirmées ?

Certaines heuristiques, contenues initialement dans le module d'apprentissage, ont pour fonction de réduire l'espace des inductions possibles et de localiser les plus désirables. Elles constituent la connaissance de base qui est formée de :

- l'information sur la pertinence des descripteurs,
- les présomptions sur la forme des descriptions optimales,
- des critères de préférence (voir §4),
- une série de règles d'inférences, heuristiques et procédures spécialisées, qui permettent un système d'apprentissage de générer d'une part les conséquences logiques d'hypothèses candidates et d'autre part de nouveaux descripteurs.

6.b) Les critères de préférence :

Ces critères servent à limiter le nombre d'hypothèses consistantes avec les instances. Parmi les mesures les plus spécifiques pour évaluer la qualité de ces hypothèses, on peut indiquer :

- la simplicité pour la compréhension humaine, mesurée, par exemple par les nombres de descripteurs et d'opérateurs utilisés,
- le degré de correspondance entre les hypothèses candidates et les instances,
- le coût de la mesure des valeurs des descripteurs utilisés,
- le coût en calcul de l'évaluation de l'hypothèse,
- la mémoire demandée pour stocker l'hypothèse,
- la quantité d'informations nécessaire pour coder l'hypothèse en utilisant les opérateurs prédéfinis.

Il est aussi possible de pondérer l'ensemble de ces aspects.

D) APPRENTISSAGE D'UN CONCEPT UNIQUE :

La finalité de cet apprentissage est de trouver, dans l'espace des descriptions possibles, le concept qui recouvre le mieux les exemples et exclut tous les contre-exemples.

Les deux hypothèses sous-jacentes à ce type de programme d'apprentissage sont d'une part que les instances d'entraînement sont toutes des exemples ou des contre-exemples d'un concept unique, et d'autre part que ce concept peut être représenté par un point dans l'espace donné des règles.

Si la 1ère hypothèse n'est pas vérifiée, on est dans le cas d'apprentissage de concepts multiples et l'approche méthodologique est différente.

Si la 2ème hypothèse n'est pas vérifiée, le programme doit être capable d'ajouter de nouveaux termes ou langage de représentation.

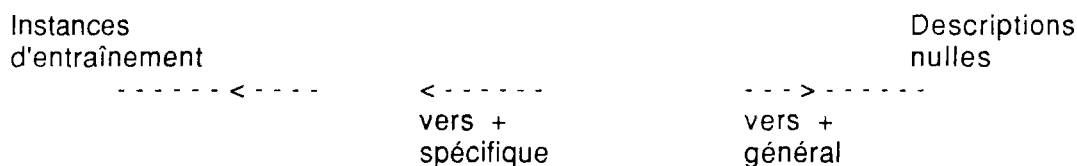
Dans ce paragraphe vont être présentés plusieurs algorithmes correspondant aux quelques méthodologies principales de ce type d'apprentissage.

1) Une recherche conduite par les données : l'algorithme d'élimination de candidats :

Le programme, décrit brièvement dans ce chapitre, explore de façon incrémentale l'espace des instances et est un exemple de modèle conduit par les données.

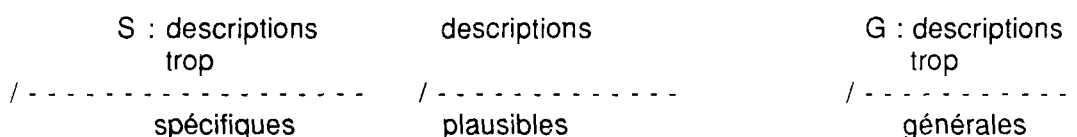
1.a) L'espace de travail :

Cet algorithme travaille dans un espace de gestion des hypothèses candidates qui peut être orienté selon 2 directions.



La description la plus générale est celle qui ne fixe aucune contrainte à aucun descripteur, c'est en ce sens qu'elle est appelée description nulle. Les descripteurs les plus spécifiques sont évidemment les instances elles-mêmes.

A un moment donné du raisonnement l'espace des règles ou hypothèses est divisé en 3 parties

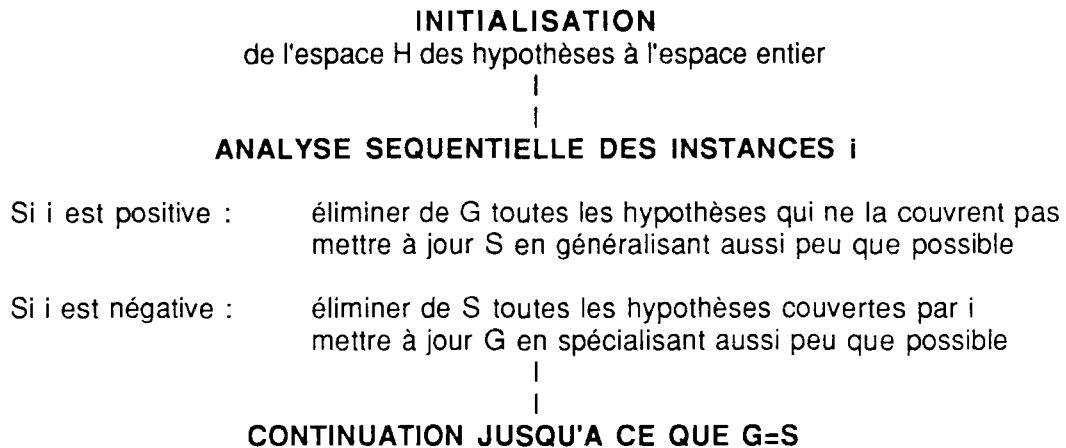


L'idée du raisonnement est de réduire progressivement l'ensemble des descriptions plausibles à une seule, en généralisant S et spécifiant G à l'aide des exemples et contre-exemples, selon l'algorithme décrit dans le paragraphe suivant.

1.b) L'algorithme

L'ensemble G désigne l'ensemble des descriptions les plus générales et ne contient initialement que la description nulle.

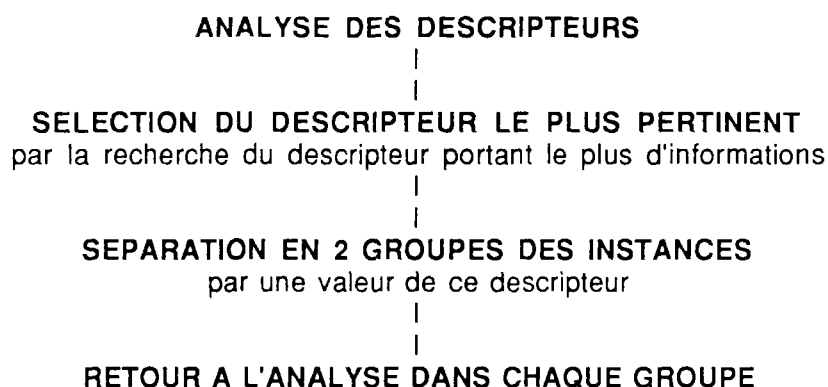
L'ensemble S est formé des descriptions les plus spécifiques, et est donc constitué au début du problème par l'ensemble des instances.



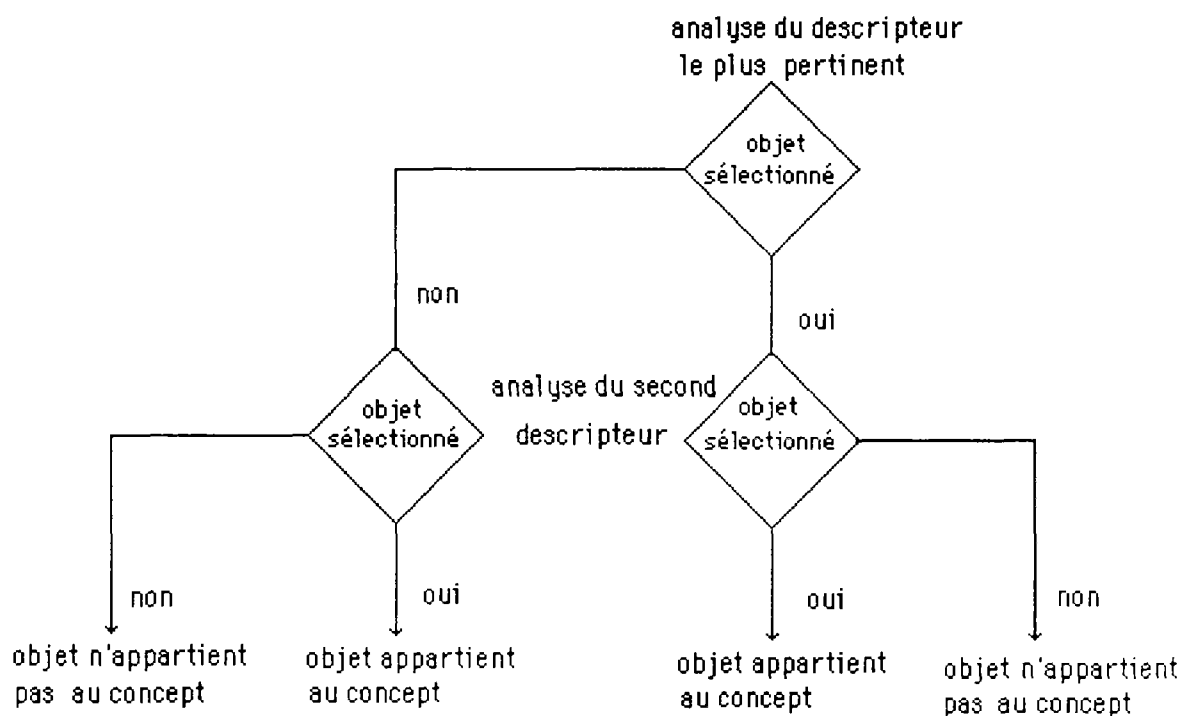
L'ensemble S représente en fait l'ensemble des conditions suffisantes pour qu'une nouvelle instance soit un exemple du concept, tandis que l'ensemble G est constitué des conditions nécessaires.

2) ID3 de QUINLAN :

L'algorithme commence avec un arbre décision vide et le construit graduellement, en ajoutant des noeuds de décision, jusqu'à ce que l'arbre classe correctement l'ensemble des instances.



L'arbre de décision ci-dessous illustre l'algorithme dans le cas très simple où seulement deux descripteurs interviennent pour partitionner l'ensemble des instances.



3) La méthodologie de l'étoile :

3.a) Le concept d'étoile :

L'étoile d'un évènement e , sous les contraintes E est l'ensemble de toutes les descriptions alternatives possibles et non redondantes de l'évènement e , qui ne transgressent pas les contraintes E .

L'étoile de l'évènement e , contre l'ensemble d'évènements E , est défini comme l'ensemble des expressions les plus générales possibles, qui couvrent l'évènement e et ne couvrent pas n'importe quel élément de E .

L'étoile peut être limitée en fixant un nombre m de descriptions sélectionnées par le critère de préférence.

3.b) Schéma de l'algorithme général pour l'apprentissage d'un concept unique :

Soient POS et NEG les ensembles représentant les exemples et les contre-exemples d'un concept. L'algorithme se présente comme suit :

1. Sélectionner au hasard un élément e dans POS
2. Générer une étoile limitée, de l'élément e contre l'ensemble des exemples négatifs NEG.

Dans le procédé de génération d'étoiles, sont appliquées :

- * les règles de généralisation (à la fois sélectives et constructives)
 - * les règles spécifiques d'une tâche
 - * les heuristiques pour générer les nouveaux descripteurs fournis par la connaissance de base du problème
 - * les définitions de concepts préalablement appris.
3. Dans l'étoile obtenue, trouver la meilleure description D selon la pondération des critères de préférence, qui a été définie.
 4. Si la description D couvre l'ensemble POS complètement, aller en 6.
 5. Autrement, réduire l'espace POS aux seuls éléments non couverts par D , et reprendre le procédé depuis l'étape 1.
 6. La disjonction de toutes les descriptions générées D est une description de concepts complète et consistante.
 7. Appliquer les règles de reformulation définies dans la connaissance de base du problème, et les règles de contraction pour obtenir l'expression la plus simple possible.

3.c) La méthode INDUCE (Michalski) :

Pour la méthode INDUCE, les étapes de la procédure sont :

1. Dans la 1ère étape, les sélecteurs individuels de l'évènement e sont mis sur une liste appelée Partial Star (PS). Cette liste peut contenir des éléments de NEG. Ces éléments de PS peuvent être vus comme des généralisations de l'élément e en appliquant de toutes les manières possibles la condition d'élimination dans la règle de généralisation.

Chaque application élimine tous les sélecteurs sauf un.

Les éléments de l'étoile partielle PS sont alors ordonnés du plus au moins préféré selon le critère de préférence.

2. La liste PS est alors étendue en additionnant de nouveaux sélecteurs en appliquant à l'élément e les règles d'inférence suivantes :
 - les règles de généralisation constructive
 - les heuristiques spécifiques au problème définies dans la connaissance de base
 - les définitions des concepts préalablement appris (pour déterminer quelles parties de e satisfont quelques concepts déjà connus).
3. Chaque nouveau sélecteur est inséré dans la position appropriée dans la liste PS selon le critère de préférence. La taille de PS est gardée dans une limite définie.
4. Les descriptions dans la liste PS sont testées pour la consistance et la complétude. Les descriptions consistantes et complètes sont retirées de la liste PS et placées dans la liste "solutions" (qui peut être limitée). Les descriptions consistantes sont aussi retirées et placées dans la liste "consistants".
5. Chaque expression dans PS est spécialisée de différentes manières en y ajoutant un sélecteur simple.
6. Chaque expression sur la liste "consistant" est généralisée en appliquant les règles d'"extension contre", de fermeture d'intervalle, et de remontée dans la hiérarchie conceptuelle.
7. Les généralisations obtenues sont rangées selon le critère de préférence. Pour obtenir une description discriminante, un mode typique d'agrégation de critères est de maximiser le nombre d'évènements couverts dans l'ensemble des possibles, et de minimiser la complexité des expressions, mesures, par exemple, par le nombre de sélecteurs qu'il contient.

Les n meilleures expressions ainsi déterminées, consistent l'étoile limitée G .

V) APPRENTISSAGE PAR EXPERIMENTATION

A) L'IDEE DE BASE :

Ce type constitue une variante de l'apprentissage à partir d'exemples qui s'applique aux systèmes réalisant des tâches séquentielles, qui sont essentiellement des tâches de planification pour lesquelles le module de réalisation doit trouver une séquence d'opérateurs menant d'un état initial à un but fixé.

Les bases de connaissances de ce type de système sont formées d'un ensemble de règles de production du type :

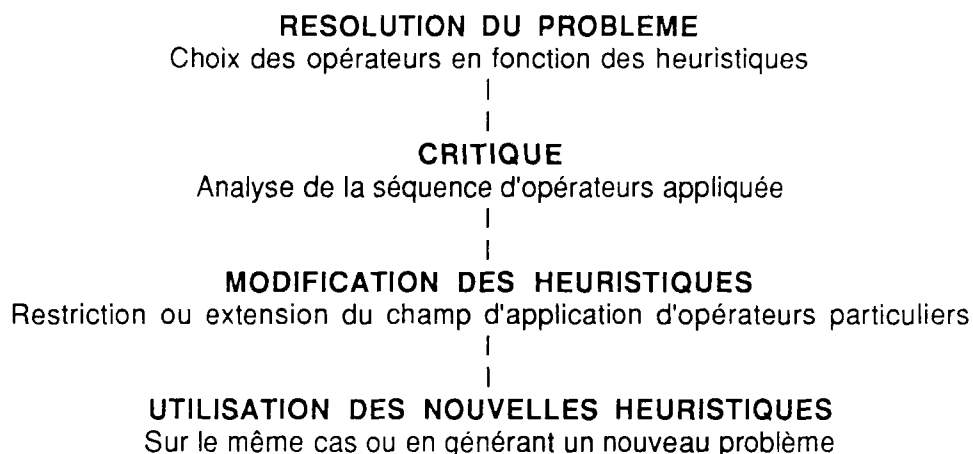
DESCRIPTIONS de SITUATIONS	=>	PRESCRIPTIONS d' ACTIONS
----------------------------------	----	--------------------------------

Le module de réalisation applique à chaque étape du raisonnement la règle dont la description de situation est la plus proche de la situation courante et dont l'action doit conduire à un nouvel état, plus proche du but visé.

Pour la plupart des programmes de ce type, la liste des opérateurs est fixée par avance et la tâche d'apprentissage consiste à déterminer dans quelles situations un opérateur ou une action doivent être appliqués. La base de connaissance contient initialement un ensemble d'heuristiques sur l'utilisation des opérateurs et c'est cet ensemble qui est progressivement enrichi par l'expérience acquise lors de la résolution des problèmes.

B) LES ETAPES D'UNE SESSION D'APPRENTISSAGE :

La boucle centrale de ce type de programme est constituée par quatre phases :



Cette présentation succincte met en évidence deux difficultés de ce type d'apprentissage. La première vient du grand nombre d'interactions possibles entre règles quand la base de connaissances est modifiée. La seconde vient du fait que la critique ne peut assigner de crédit ou de blâme aux heuristiques avant la résolution complète du problème. Aussi, s'il est facile de déterminer si la séquence entière est bonne ou mauvaise, il n'est pas évident de déterminer quelles heuristiques ont contribué au succès ou à l'échec de la solution.

Quelques exemples vont maintenant illustrer des aspects particuliers de l'apprentissage par expérimentation.

C) LE PROGRAMME D'INTEGRATION SYMBOLIQUE LEX :

1) La fonction du programme :

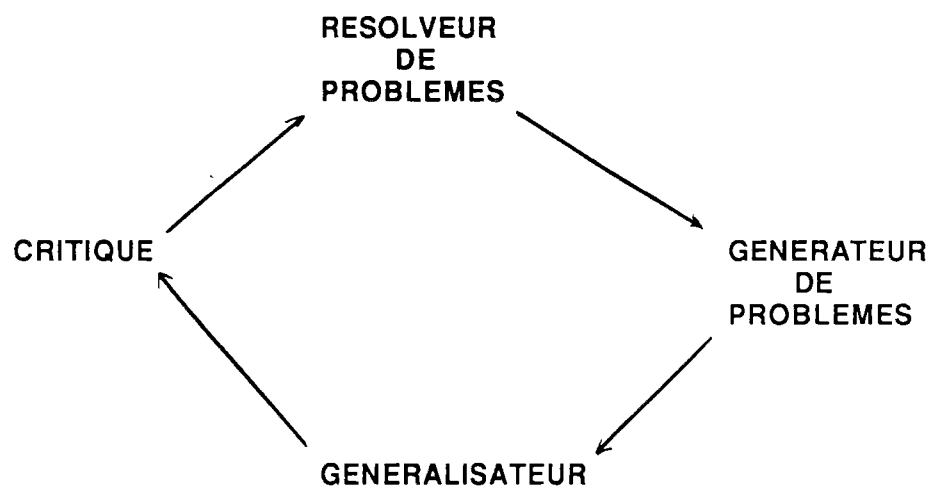
LEX apprend à résoudre les problèmes d'intégration symbolique en enrichissant un ensemble d'heuristiques indiquant dans quelle situations différents opérateurs devraient être utilisés. LEX dispose d'un ensemble n d'opérateurs OP_i ($1 < i < n$) et le but de LEX est de développer des règles de production de la forme :

$\langle \text{Classe de situations} \rangle \Rightarrow \text{utilisation de l'opérateur } OP_i$

LEX définit progressivement les concepts de situations en recueillant et archivant toutes les instances positives et négatives pour chaque opérateur.

2) L'architecture du programme

LEX est organisé autour de quatre modules, travaillant séquentiellement en boucle.



Le résolveur correspond au module de réalisation, l'ensemble de critique et du généralisateur assurent la fonction d'apprentissage, et le générateur de problème va organiser la validation des heuristiques définie lors de l'apprentissage, en construisant des cas appropriés.

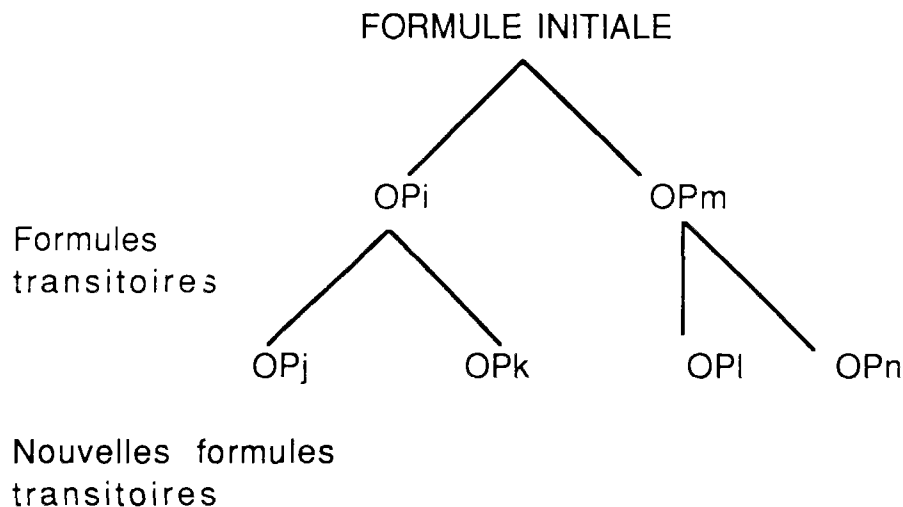
La construction de l'arbre s'arrête quand l'intégration est résolue ou après un nombre d'essais déterminé, qui indique l'échec du problème.

En cas de succès la trace du raisonnement est gardée.

3) Le rôle du résolveur de problème :

Il recherche les applications possibles d'opérateurs dans les tentatives de résolution d'un problème d'intégration donné. Initialement, la recherche est aveugle, mais au fur et à mesure de la définition des heuristiques délimitant le champ d'application des opérateurs, le choix des opérateurs, à une étape donnée de l'intégration, est de plus en plus restreint.

Le résolveur en construisant une arborescence



Après chaque application d'opérateur, l'arbre est évalué avant de continuer l'exploration d'une branche particulière. A chaque noeud, le programme associe la somme des coûts d'application des opérateurs depuis la racine, puis détermine, en suivant les indications des heuristiques, l'opérateur le plus prometteur auquel est associé un poids.

Une pondération est alors effectuée entre le coût actuel de chaque noeud et l'intérêt de son ouverture, pour déterminer la branche de l'arbre à créer et à explorer en priorité.

4) Le rôle du critique

Il analyse les traces de chaque problème résolu en assignant un crédit ou un blâme à chaque décision individuelle repérée dans le raisonnement.

Si une solution a pu être trouvée par 2 chemins différents, le critique considèrera comme décision positive chaque étape sur le chemin du coût minimum, et comme décision négative chaque étape allongeant d'un coefficient donné le coût de la trace minimale.

5) Le généralisateur

Il applique l'algorithme d'élimination de candidats pour mieux cerner l'espace d'application de chacun des opérateurs, à partir de l'analyse de l'ensemble des décisions positives et négatives définie par le critique.

A chaque étape, il essaie de définir le concept décrivant les situations correspondant aux champs d'application des opérateurs.

6) Le générateur de problème

Le dernier module de LEX analyse les concepts correspondant à chaque opérateur, et détermine le problème pertinent pour affiner un concept particulier, en le généralisant ou en le spécialisant.

Ainsi, le déclenchement d'une nouvelle session d'apprentissage aura pour objet de restreindre ou étendre le champ d'application d'un opérateur particulier.

D) APPRENTISSAGE PAR GENERATION ET TEST D'HYPOTHESES

1/ Le programme DENDRAL

La finalité du programme a été décrite dans le chapitre 2.

2/ L'introduction d'un module d'apprentissage (META-DENDRAL) :

a/ La finalité du module d'apprentissage

L'objectif de ce module est la découverte automatique des règles de fragmentation moléculaire utilisées dans le programme de recherche de formules.

Ce module est décomposé en trois parties, chacune correspondant aux trois étapes suivantes :

- l'examen des exemples avec leurs spectres pour déterminer quelles liaisons cassent (programme INTSUM)
- la recherche des structures communes autour des liaisons rompues et la proposition de règles du type (structures) => (fragments) (programme RULEGEN)
- affinage et la modification des règles proposées, pour améliorer l'explication des données observées. (programme RULEMOD)

b/ L'examen des exemples

Ce programme travaille avec un ensemble d'exemples constitués de molécules avec leur formule et leur spectre de masse.

Il commence par simuler une spectrographie de masse de chaque molécule, en partant sur l'hypothèse d'une fragmentation complète. Le spectre artificiel produit contient alors beaucoup plus de pics qu'un spectre réel.

Pour chaque pic du spectre artificiel, le programme enregistre la règle de fragmentation qui en est à l'origine. Puis il compare ensuite les spectres artificiel et réel d'une molécule, et infère que chaque pic commun résulte de la règle de fragmentation qui a généré le pic dans la règle spectre artificiel.

A ce niveau du raisonnement, le programme a mis en évidence des liaisons entre des pics et des emplacements de liaisons rompues. La tâche suivante est d'analyser la nature des liaisons cassées et leur environnement. C'est le rôle du programme RULEGEN.

c/ La création de règles

Cette partie recherche les corrélations entre les ruptures observées et les environnements de liaisons chimiques autour de la rupture observée.

La création d'une règle se fait pas spécialisation de la règle la plus générale possible, décrite comme $R+R$, où R représente n'importe quel atome ou groupe d'atome et $+$ la liaison rompue.

RULEGEN spécifie en ajoutant des traits particuliers, c'est-à-dire en décrivant de plus en plus précisément le groupement R .

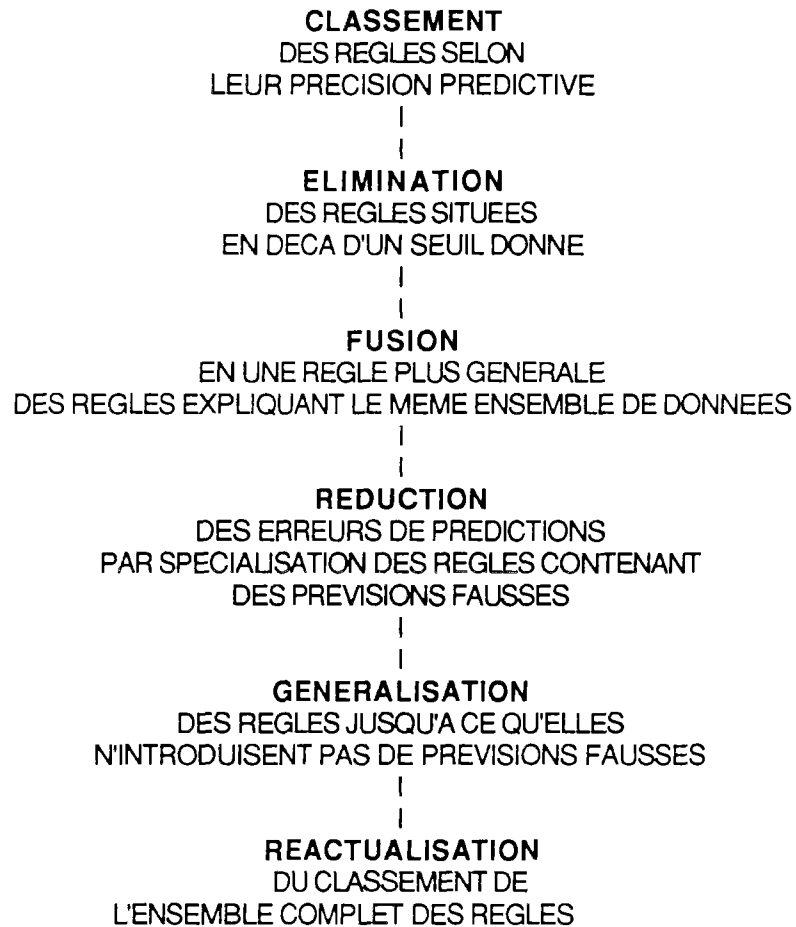
RULEGEN commence donc par le parent $R+R$ et génère des générations successives d'enfants, définissant ainsi implicitement un arbre généalogique, où à chaque étape est introduit un nouvel atome.

La structure arborescente définit l'espace dans lequel META-DENDRAL conduit sa recherche, dirigée par un modèle, composé d'un ensemble d'hypothèses prioritaires sur la génération des spécifications pour un parent donné.

d/ Modification et affinage de règles

La troisième partie, RULEMOD, a pour fonction de modifier et d'affiner les règles, pour mieux les caler sur les données observées.

RULEMOD examine à la fois les prévisions correctes et les prévisions incorrectes obtenues par application des règles, et modifie itérativement les règles selon la boucle suivante :



VI) L'APPRENTISSAGE INDUCTIF A PARTIR D'OBSERVATIONS :

A) L'IDEE GENERALE :

Dans l'apprentissage inductif à partir d'exemples, la tâche d'apprentissage consistait à décrire un concept préalablement défini en recherchant à travers le jeu des exemples et des contre-exemples sa meilleure description. Dans les apprentissages à partir d'observations, le système reçoit des informations de bas-niveau qu'il doit analyser pour les structurer ou découvrir des relations. Les tâches d'apprentissage de ces programmes peuvent donc être formalisées et se rattacher à un des types suivants.

1) La formulation d'une théorie :

Formuler une théorie signifie le plus souvent découvrir une relation constante entre les différentes données, comme par exemple une relation entre les différents attributs d'un objet. Par exemple, l'objet GAZ est caractérisable par les 3 attributs que sont sa pression, son volume, et sa température. La théorie des gaz parfaits indique que ces 3 attributs ne sont pas indépendants mais reliés par la relation $PV/T = Cte$.

Le programme d'apprentissage BACON, présenté dans le paragraphe B a pour finalité la découverte de telles relations.

2) La recherche de structures dans les données observées

2.a) De l'analyse de données à la notion de regroupement conceptuel :

Les premières recherches sur les classifications automatiques d'observations ont été menées par les statisticiens qui ont développé des algorithmes d'analyses de données. Leurs méthodes utilisent des représentations "attribut-valeur" pour les objets, vus comme des points dans un espace à n dimensions. La similitude entre 2 objets ou 2 regroupements est mesurée par leur distance dans cet espace et ces méthodes recherchent la classification qui maximise la similitude interne et minimise celle entre les groupements, en utilisant une fonction numérique basée sur une métrique euclidienne.

L'approche de l'intelligence artificielle dans la formation de taxinomies est différente dans le sens qu'elle introduit une approche heuristique en sélectionnant les dimensions pertinentes sur lesquelles sera basée une classification et en introduisant la notion de regroupement conceptuel.

2.b) Exemples de concepts non obtenus par une métrique euclidienne :

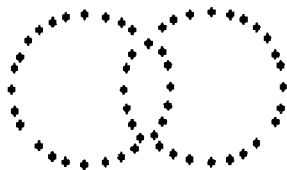


figure A

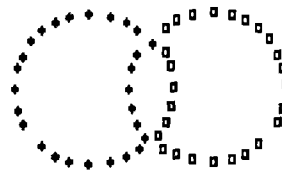


figure B

La figure de gauche représente un ensemble de points dans un plan, repérés par leurs coordonnées cartésiennes. La figure de droite représente les 2 ensembles de points obtenus par un algorithme classique de regroupement basé sur la minimisation de la distance euclidienne des points. Or une simple visualisation du premier schéma indique que ces points forment en fait 2 cercles qui s'enchevêtrent.

Le rassemblement de ces points en deux cercles demande que leur regroupement ne soit pas basé uniquement sur leur proximité, mais soit aussi fonction de leur appartenance à la définition d'un même concept.

En se basant sur de tels exemples, Michalski et Stepp (Machine Learning) ont critiqué les méthodes numériques de classifications parce que ce sont les objets eux-mêmes qui sont utilisés dans l'évaluation des regroupements concurrents, à chaque étape de la formation de la taxinomie. Michalski et Stepp ont pour idée d'introduire des critères portant sur la qualité des schémas de regroupement, en donnant par exemple une préférence aux descriptions simples ou à celles qui ont le plus grand pouvoir prédictif, en guidant la formation des classifications en leur donnant un but ou en définissant la notion d'appartenance à un concept, définie comme la cohésion conceptuelle.

2.c) La notion de cohésion conceptuelle :

Dans les analyses de données conventionnelles, la similitude entre 2 objets est caractérisée par un simple nombre, correspondant à la valeur de la fonction de similitude appliquée aux différents descripteurs des objets. Les fonctions de similitude sont indépendantes du contexte et incapables d'extraire les propriétés globalisantes des regroupements d'objets, c'est-à-dire les propriétés qui caractérisent un ensemble comme un tout et ne sont pas déductibles de ses éléments individuels, comme les 2 cercles dans l'exemple du paragraphe précédent.

Pour détecter de telles propriétés, caractérisant l'appartenance de l'objet à une configuration particulière, le système doit avoir la capacité de reconnaître les configurations d'objets représentant certains concepts globaux.

En conséquence, la similitude entre 2 objets ne dépend pas seulement de leurs propriétés et des propriétés des objets à proximité, mais aussi de l'ensemble des concepts disponibles pour décrire A et B ensemble.

2.d) La construction d'une taxinomie :

Des regroupements de niveaux supérieurs peuvent aussi être effectués et leur hiérarchisation successive conduit à la construction d'une taxinomie. Ce type de programme construisant une classification à partir d'une collection d'objets, présentés comme autant d'observations, est souvent appelé dans la littérature apprentissage par détection de similarité, et le programme CLUSTER décrit dans le paragraphe C en est un bon exemple.

B) APPRENTISSAGE PAR DECOUVERTE DE LOIS : LE PROGRAMME BACON :

1) La finalité du programme :

Le programme BACON cherche à reproduire les étapes de la découverte d'une loi de variation, comme les lois d'Ohm pour les circuits électriques, les lois des gaz parfaits ou les lois de Kepler pour le mouvement des planètes autour du soleil.

Ce programme manipule des attributs numériques appartenant à des objets spécifiés. L'objet dipôle électrique est caractérisé par 2 attributs, l'intensité du courant qui le traverse et la tension à ses bornes. L'objet planète possède 2 attributs, la distance au soleil et sa période de révolution.

2) La méthodologie suivie par BACON :

2.a) L'idée générale :

Le programme BACON utilise des heuristiques pour créer de nouvelles variables, appelées les termes théoriques, à partir d'une combinaison des attributs initiaux des objets. Bacon cherche à découvrir le terme théorique constant pour l'ensemble des objets constituant les observations.

2.b) Quelques exemples d'heuristiques :

Certaines heuristiques indiquent la combinaison de variables à étudier en priorité en fonction des évolutions comparées des valeurs des attributs.

Par exemple, dans les résistances électriques, l'attribut tension U augmente en même temps que l'attribut intensité I , aussi le programme définira comme premier terme théorique le rapport U/I .

Quand le problème se complique et que plusieurs termes théoriques ont été créés, et un autre jeu d'heuristiques indique ce qui doit être prioritairement combiné.

3) Les limites du programme BACON :

Ce programme a beaucoup de difficultés à travailler avec des données bruitées, et quand le nombre de variables augmente, il rencontre des difficultés à extraire les variables pertinentes et à construire les bons termes théoriques.

D'autre part, la découverte de lois qualitatives n'est qu'un des aspects de la découverte scientifique. D'autres programmes, comme GLAUBER ont pour objet la génération de lois qualitatives.

Enfin, certains auteurs, comme BRADSHAW, LANGLEY, SIMON ET ZYTKOW (in Machine Learning II) essaient de combiner les deux approches pour construire des programmes d'apprentissage capables de construire des modèles structurels.

C) APPRENTISSAGE INDUCTIF PAR DETECTION DE SIMILITUDES :

1) L'idée de regroupement conceptuel basé sur les attributs (AC3) :

1.a) La finalité du programme :

Ce type de programme doit structurer une collection d'observations formant un ensemble d'objets caractérisés chacun par une liste d'attributs nominaux, linéaires ou structurés.

Les premiers travaux sur les classifications de Michalski et Stepp, indiquent qu'une configuration d'objets forme une classe si, et seulement si, elle peut être décrite par un concept conjonctif impliquant des relations sur les attributs des objets. Ceci se formalise ainsi :

étant donné :

- un ensemble d'objets (physiques ou abstraits),
- un ensemble d'attributs caractérisant les objets,
- une connaissance de base qui inclue les contraintes du problème, les propriétés des attributs, les règles d'inférence pour générer de nouveaux attributs, et un critère pour évaluer la qualité des classifications candidates.

trouver :

- une hiérarchie de classes d'objets et leurs descriptions sous forme d'états conjonctifs.

Les sous-classes descendantes de n'importe quelle classe parente doivent logiquement avoir des descriptions distinctes et maximiser un critère de préférence de regroupement.

1.b) Les critères de mesure de qualité des regroupements :

A plusieurs étapes de la classification, le programme devra choisir entre des regroupements concurrents. Ce choix est basé sur une mesure combinée de plusieurs critères, qui dans l'algorithme CLUSTER de Michalski prennent en compte :

- le degré de correspondance entre le regroupement et les observations,
- la simplicité des descriptions de regroupement,
- la différence inter-regroupements,
- l'index de discrimination,
- la réduction de dimensionnalité.

Le degré de correspondance est apprécié par l'étude des dispersions des différentes descriptions des regroupements.

La simplicité est liée au nombre total de sélecteurs dans toutes les descriptions.

La différence inter-regroupement est calculée à partir des degrés de disjonction entre chaque paire de descriptions de regroupements.

L'index de discrimination correspond au nombre de variables ayant différentes valeurs dans chaque description de regroupement.

La réduction de dimensionnalité dépend du nombre minimum de variables nécessaires pour distinguer toutes les descriptions dans un regroupement.

1.c) Les deux phases de la méthode de regroupement conceptuel :

Le regroupement conceptuel d'un ensemble de données commence par une phase de regroupement et se continue par une phase de construction de hiérarchies.

Pendant la première phase, les objets sont rangés en classes de telle sorte que le regroupement obtenu maximise le critère de préférence donné et dépendant du contexte.

L'algorithme sélectionne alternativement un ensemble d'objets "semences" pour guider les inférences inductives produisant des descriptions généralisées, mais mutuellement disjointes de classes d'objets. Ce procédé assure que chaque objet semence est placé dans une classe distincte.

Chaque description est aussi générale que possible, de telle sorte qu'elle couvre la semence donnée, mais aucune autre. L'algorithme est détaillé au paragraphe suivant.

La phase de construction de hiérarchie commence avec la construction de classification de premier niveau de tous les objets, puis continue par une construction récursive de classifications pour chaque groupe d'objets.

1.d) L'algorithme de regroupement conceptuel :

1.d.i) Les notions de base de l'algorithme :

La distance syntaxique entre 2 objets e_1 et e_2 :

Elle est définie comme la somme des distances syntaxiques entre les valeurs de chaque descripteur des objets e_1 et e_2

La distance entre 2 valeurs de descripteurs dépend du type du descripteur.

S'il est nominal, la distance syntaxique est nulle si les valeurs sont identiques et égale à 1 si elles sont différentes.

S'il est linéaire, la distance est égale au rapport de la différence absolue entre les valeurs sur l'échelle de variation du descripteur.

S'il est structuré, la distance dépend de la place des valeurs dans la hiérarchie de généralisation.

La notion de recouvrement :

Une description recouvre des objets quand elle reconnaît ces objets et rejette les autres. Un recouvrement est formé de l'ensemble des valeurs dans la hiérarchie des généralisations.

La notion de dispersion :

Soient E un ensemble d'objets et E_1 le sous-ensemble des objets de E que l'on désire regrouper ensemble. Soit D une description et $n(D)$ et $n_1(D)$ les nombres respectifs d'objets de E et de E_1 couverts par D .

La dispersion d de D dans E relative à E_1 est définie par la relation :

$$d = 1 - n_1(D)/n(D)$$

Si $n_1(D) = n(D)$, tous les éléments couverts par D sont dans E_1 et leur dispersion est donc nulle.

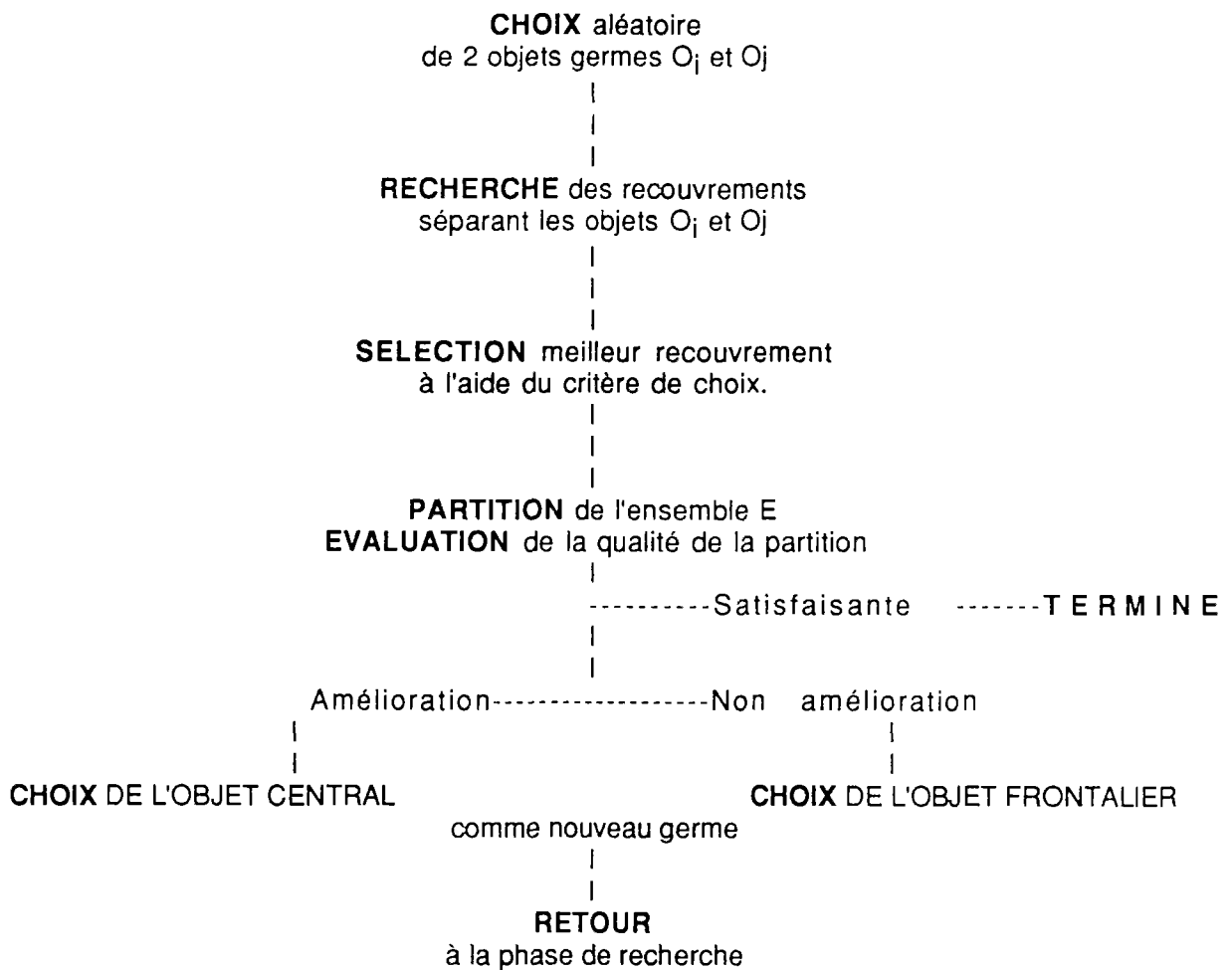
Si $n_1(D) = 0$, le regroupement E_1 n'est pas adapté à la dispersion et la dispersion est donc maximale.

La notion de centralité et de frontière :

On peut calculer la somme $\sum O_i$ des distances dO_{ij} de chaque objet à tous les autres objets O_j de l'ensemble E .

L'objet pour lequel cette somme $\sum O_i$ est minimale est dit central et inversement celui pour lequel cette somme est maximale est dit frontalier.

1.d.ii) Les notions de base de l'algorithme :



Cet algorithme utilise une distance prenant en compte les connaissances sur les relations symboliques entre les objets, et recherche des regroupements stables par rapport aux notions d'objet le plus central ou le plus excentré.

1.d.3) Les limites de l'algorithme :

Il ne permet pas de générer de nouveaux descripteurs, ni de prendre en compte les buts et finalités associées aux problèmes de classifications, ce qui a été réalisé dans la version suivante de CLUSTER.

2) La seconde version de CLUSTER :

Toujours réalisée par STEPPet MICHALSKI, elle a été présentée dans l'ouvrage MACHINE LEARNING, tome II.

2.a) L'extension du travail précédent :

Cette recherche prolonge le travail précédent dans 3 directions :

- * les objets et les classes sont décrites par des descriptions structurales qui sont exprimées dans un "calcul de prédicats annotés", qui est un calcul de prédicats avec des opérateurs additionnels.
- * la connaissance de base comprend des règles d'inférence pour extraire des concepts descriptifs de haut niveau à partir des concepts de bas niveau initialement fournis.
- * le système est approvisionné avec un but général de la classification qui fournit les moyens d'identifier les descripteurs pertinents et les règles d'inférences pour obtenir de nouveaux descripteurs.

Dans cette méthode, la connaissance de base consiste en un réseau de buts de classifications, de règles d'inférences et d'heuristiques pour obtenir de nouveaux descripteurs, définition de domaines attributs et de types, et la classification de critères de préférence.

Le réseau de buts, appelé le réseau de dépendances des buts, est utilisé pour guider la recherche des descripteurs pertinents et des règles d'inférences.

2.b) La connaissance de base contenue dans le programme :

Elle est composée d'une connaissance générale et d'une connaissance spécifique au domaine. La connaissance générale indique les contraintes fondamentales des classifications et les critères spécifiant les préférences pour le regroupement des objets.

La connaissance spécifique du domaine est formée des règles d'inférences construisant les nouveaux descripteurs et le réseau de dépendance des buts, pour indiquer quels descripteurs (attributs fonctions, ou prédicats) sont pertinents par rapport au but de la classification.

2.c) Les règles de construction de nouveaux descripteurs :

De nouveaux descripteurs peuvent être obtenus de différentes manières.

Les descripteurs déduits par inférence logique sont des prédicats et des fonctions obtenus en appliquant des règles d'inférences aux descriptions initiales des objets. Par exemple, pour des objets caractérisés par leur masse M et leur volume V , l'application de la règle

" Si $M/V < \rho_{eau}$ ALORS objet_insumersible = Vrai "

définit un nouveau descripteur partitionnant l'ensemble des objets en 2 classes.

Des descripteurs sont également obtenus par des calculs spéciaux, expérimentations et appareils.

Par exemple, le programme peut compter le nombre de parties composant un objet, et définir ainsi un nouveau descripteur utile pour certains types de classifications.

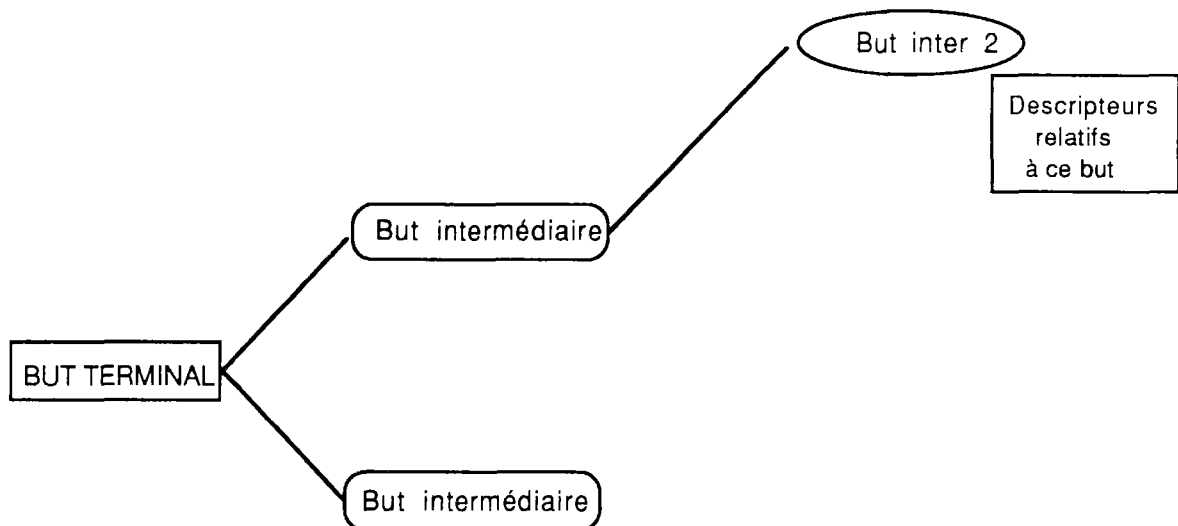
2.d) Le réseau de dépendances de buts :

2.d.i) La finalité de ce réseau :

Ce réseau va guider et structurer les différentes étapes de la classification des observations en fonction d'un but qui a été défini.

Dans l'ensemble des étapes de la classification, les descripteurs qui n'apparaissent pas dans le réseau de dépendances ne sont pas pris en compte, ce qui réduit l'espace de recherche aux combinaisons des seuls descripteurs pertinents, qu'ils soient initiaux ou construits en cours de session.

2.d.ii) L'architecture du réseau :



Ce réseau est composé d'une structure arborescente dont la racine est le but final et dont les différentes branches sont des buts intermédiaires. Chaque noeud de l'arbre est complété par deux types d'indications :

- les descripteurs pertinents pour les buts intermédiaires,
- les relations imposées entre les différents descripteurs.

3) Deux méthodes pour construire des classifications d'objets structurés :

Les procédures de classification prennent appui, d'une part sur des mesures constantes de la qualité des classifications envisagées, d'autre part sur l'utilisation de la connaissance de base, et plus particulièrement, du réseau de dépendances des buts.

3.a) Diriger le processus en mesurant la qualité des classifications :

Une fois définis les différents critères spécifiant les propriétés souhaitées d'une classification, ceux-ci sont agrégés en un seul critère appelé le critère fonctionnel d'évaluation lexicographique aux tolérances.

Ce critère fonctionnel est constitué d'une séquence hiérarchisée de critères élémentaires avec leurs tolérances. Ces tolérances indiquent jusqu'à quelle extension, différentes solutions sont considérées comme équivalentes et ne doivent donc pas être distinguées du point de vue de ce critère et seront donc comparées sur la base du critère suivant dans la hiérarchie.

3.b) L'utilisation de la connaissance de base

La construction d'une classification pleine de signification revient à trouver de bons attributs de classification, c'est à dire des attributs de haut niveau pour définir des classes.

Le réseau de dépendances des buts est alors scruté pour trouver les interactions entre les buts d'une classification et les descripteurs potentiels.

La connaissance de base rend le système capable de réaliser une chaîne d'inférences pour déduire les nouveaux descripteurs et leurs valeurs, qui seront inclus dans les descriptions des objets.

Les paragraphes suivants présentent deux méthodes qui utilisent de façon différente le critère fonctionnel et la connaissance de base pour former les classes et leur concept.

3.c) Formation de concepts par la méthode de discrimination répétée :

Cette méthode résout le problème de formation de concepts, dans ce cas la construction d'une classification, par une séquence d'étapes contrôlées d'acquisition de concepts, basée sur l'apprentissage à partir d'exemples.

Comme le nombre k de classes à construire n'est pas spécifié à l'avance, l'algorithme construira des regroupements de 2 à 7 classes, sélectionnera le meilleur regroupement pour chaque nombre k , et confrontera ensuite les regroupements sélectionnés.

Préalablement à l'étape de recherche de la classification optimale, le programme recherche la généralisation caractéristique optimisée de la collection entière des observations. Cette généralisation caractéristique exprime la sous-structure commune à tous les objets. Sa définition décompose l'objet en un sous-ensemble de variables libres (indépendantes de l'objet), et ses parties spécifiques. Cette portion de la structure de chaque objet décrite par la généralisation caractéristique est appelée le noyau de l'objet, qui n'est pas prise compte dans l'étape de classification.

L'algorithme d'acquisition de concepts pour un nombre k de classes est celui d'INDUCE.

1. K objets "semence" sont sélectionnés au hasard et traités comme les représentants individuels de K classes imaginaires.
2. L'algorithme génère alors les descriptions de chaque semence qui sont les plus générales et ne couvrent aucune autre semence.
3. Les descriptions obtenues sont alors utilisées pour déterminer l'objet central le plus représentatif dans chaque classe nouvellement formée
4. Les objets représentatifs sont alors utilisés comme nouvelles semences pour l'itération suivante.
5. Cette répétition cesse quand des itérations consécutives convergent vers une solution stable ou quand un nombre donné d'itérations est atteint sans qu'il y ait amélioration de la classification.

Cette approche demande que cet algorithme soit repris autant de fois qu'il y a de nombres de classes possibles.

3.d) Formation de concept en trouvant des attributs classifiants

Cette méthode de classification recherche un ou plusieurs attributs classifiants dont les ensembles de valeurs peuvent être fractionnés en champs définissant des classes individuelles.

L'aspect important de cette approche est que l'attribut classifiant peut être obtenu, à partir des attributs initiaux, à travers une chaîne d'inférences dirigées par le but donné à la classification.

Les attributs classifiants recherchés sont ceux qui conduisent aux classes d'objets correspondant aux indications de pertinence attachées aux noeuds du réseau pour les buts intermédiaires de la classification.

L'intérêt présumé qu'un descripteur donné serve comme attribut classifiant est déterminé d'une part en consultant le réseau de dépendances des buts, et d'autre part en considérant combien d'autres descripteurs sont déductibles à partir de lui. Les descripteurs de haut niveau hiérarchique, contenant implicitement le plus d'attributs, sont préférés.

Il y a deux procédures fondamentales qui opèrent alternativement pour générer la classification.

La première (**SEARCH**) recherche les attributs classifiants dont l'ensemble de valeurs peut être partitionné pour former des classes telles que la classification produite soit la meilleure selon le critère fonctionnel. Cette procédure de recherche est activable de 2 manières.

Quand le nombre k de classes est connu par avance, elle recherche les attributs possédant au moins k valeurs dans les descriptions des objets à classifier. Les attributs avec un nombre de valeurs inférieur à k ne sont pas pris en considération.

Pour les attributs ayant des ensembles de valeurs observées supérieures à k , le choix de l'application des sous-ensembles de valeurs vers les classes résulte de l'évaluation du critère fonctionnel sur l'ensemble des applications possibles.

Quand le nombre de classes est inconnu, la méthode décrite est reproduite pour le spectre des valeurs possibles de k , et la sélection est toujours basée sur l'évaluation du critère fonctionnel.

La seconde procédure (**GENERATE**) génère de nouveaux descripteurs par une chaîne d'inférences utilisant les deux formes de règles de connaissance de base, c'est-à-dire les règles implicatives logiques et les règles arithmétiques.

Certaines heuristiques sont utilisables pour guider la construction de nouveaux descripteurs. Par exemple, l'analyse comparative des variations pour des descripteurs numériques oriente le choix de l'opérateur de combinaison vers le multiplicateur ou le diviseur. D'autres heuristiques sélectionnent les mécanismes d'héritage à utiliser en priorité.

Les descripteurs ainsi inférés sont ordonnés par évaluation de leur pertinence vis-à-vis du réseau de dépendances des buts et des buts de la classification.

VIII) APPRENTISSAGE PAR ANALOGIE

A) LA SIGNIFICATION DU TERME ANALOGIE EN APPRENTISSAGE :

L'intuition centrale sur l'analogie laisse entendre qu'une structure d'un domaine peut être appliquée dans un autre domaine avec des objets arbitrairement différents. C'est le premier sens donné à l'analogie, qui peut être compris comme un transfert d'informations, par mise en relation d'objets ou groupes d'objets. C'est en ce sens qu'on parlera d'analogie relationnelle.

Dans les travaux de Gentner et Forbus, une analogie est définie comme une comparaison dans laquelle des structures relationnelles, mais peu ou pas d'attributs d'objets, sont apparentées d'une source à une cible. Les relations apparentées sont définies par l'existence d'un ensemble de relations contraignantes de haut niveau entre les objets.

Les correspondances entre objets du domaine source et ceux du domaine cible sont déterminés par le rôle des objets dans la structure relationnelle, et non par quelque similitude intrinsèque entre les objets eux-mêmes. Par exemple, en thermodynamique, l'attribut température d'un corps joue le même rôle que la pression en statique des fluides, car le transfert de chaleur se fait toujours du corps chaud vers le corps froid, de même que le liquide se déplacera des hautes vers les basses pressions.

En ce sens, l'analogie est différente de la similitude littérale où à la fois les attributs et les relations sont apparentables de la source à la cible. Ainsi l'eau et l'alcool vont pouvoir être caractérisés par les mêmes jeux d'attributs et les mêmes relations entre attributs (pression, température, volume,...).

Si le domaine source ne contient que des relations abstraites, et donc pas d'attributs d'objets les apparentements de relations seront définis comme des abstractions. L'eau et l'alcool de l'exemple précédents sont tous des liquides, et suivront les lois générales de comportement des liquides.

Travaillant également sur l'apprentissage par analogie, Berstein a développé un programme de formation de concepts par raisonnement analogique incrémental, qui, en maintenant en mémoire des enregistrements liés aux analogies effectuées, ajuste par étape les concepts en fonction des exemples et des nouvelles spécifications.

L'apprentissage par analogie recouvre également la résolution de problème par analogie, où l'expérience passée est exploitée en reprenant les solutions définies dans des situations proches, c'est-à-dire les séquences d'opérateurs utilisées. L'analyse des solutions des problèmes similaires antérieurs est objet des travaux de Carbonnell, qui a introduit les termes d'analogie transformationnelle et dérivationnelle.

L'analogie transformationnelle est définie comme une méthode de résolution dans laquelle la solution d'un problème similaire est transformée de façon incrémentale en une solution du nouveau problème. Dans l'analogie dérivationnelle, les différentes étapes de résolution des problèmes similaires sont reprises et modifiées pour résoudre le nouveau problème.

B) APPRENTISSAGE PAR ANALOGIE RELATIONNELLE :

1) Aspects principaux de ce type d'apprentissage :

Ce type d'algorithme a pour finalité d'exploiter une information donnée au système, spécifiant l'analogie, (par exemple : l'objet A est analogue à l'objet B, ou bien les objets A et B ont des relations du même type que les objets C et D), .

L'algorithme devra donc être capable d'identifier dans un premier temps les informations susceptibles d'être transférées, puis dans un second temps, de contrôler la pertinence du transfert d'informations par analogie.

Dans d'autres cas, non abordés ici, l'environnement ne précise pas les objets sur lesquels porte l'analogie, et le programme doit commencer par une phase d'identification des relations possibles et pertinentes.

Forbus et Gentner ont développé la théorie des processus qualitatifs et la théorie de l'application structurelle pour servir de support à ce type d'apprentissage.

2) La théorie des processus qualitatifs :

Cette théorie modélise une situation physique par une collection d'objets et leurs relations. Les propriétés quantitatives d'un objet sont représentées par une valeur et un sens de variation.

Dans cette théorie, les valeurs numériques ne sont jamais utilisées, mais servent à définir l'espace des quantités, formé de l'ensemble des inégalités entre les valeurs quantitatives.

Un processus physique est caractérisé par 5 éléments :

- les objets individuels concernés,
- les préconditions de son activation,
- les conditions sur les inégalités dans l'espace des quantités,
- les relations,
- les influences.

Par exemple, le déplacement d'un liquide entre deux récipients est caractérisé par :

- les récipients A et B et le tuyau les reliant,
- une précondition : le tuyau doit être plein de liquide,
- une condition d'inégalité : la pression du liquide en A > pression du liquide en B,
- une relation : le débit est proportionnel à la différence de pression,
- des influences : le volume de liquide augmente dans B et diminue dans A.

Les préconditions, conditions et relations forment un ensemble cohérent appelé plus loin structure.

3) La théorie de l'application des structures :

Lorsqu'une analogie est spécifiée entre un domaine source et un domaine cible, la recherche des informations à transférer et leur transfert demandent plusieurs phases.

a) Identification des concepts contenus dans les objets source et cible.

b) Evaluation de la pertinence de l'information donnée au système, par exemple en contrôlant l'appartenance des objets des domaines source et cible à des classes compatibles au regard de

l'information transférée. Il s'agit de voir si l'analogie spécifiée entre les objets concernés est plausible. Cette phase, qui n'a rien de trivial, est très importante.

c) Création d'une structure de correspondance à partir du domaine source vers le domaine cible de façon à ce que les sous-concepts des deux domaines puissent être tous mis en parallèle. La construction de la structure de correspondance s'appuie sur le principe de systématisme. Ce principe indique que les prédicats à exporter dans le domaine cible forment vraisemblablement un ensemble cohérent de relations et de conditions contraignant ces relations. Cette phase se décompose comme suit :

- recherche des structures de correspondance pré-existantes associées au domaine source. (Cette structure peut avoir été construite lors d'une analogie précédente).
- si une telle structure est trouvée, vérifier qu'elle est appropriée au cas courant.
- si aucune structure n'est trouvée, alors déclencher l'algorithme d'identification, qui recherche l'ensemble des similitudes entre les deux domaines. Cette phase de recherche peut être guidée par des heuristiques extrayant les concepts "forts" du domaine source ou se focalisant sur les types de connaissance les plus chargées de signification.

d) Contrôle du transfert : une fois construite la structure et que les éléments des deux domaines correspondent un à un, la connaissance du domaine source peut être ajoutée au domaine cible, après vérification que les inférences analogiques sont bien compatibles avec le domaine cible.

4) Application de structures et apprentissage :

L'intérêt d'une comparaison et des applications qui en découlent dépend de l'accessibilité et de l'utilité des analogies spécifiées. L'accessibilité est définie comme la vraisemblance que la bonne identification sera faite sur les relations.

L'utilité correspond à la quantité d'informations déductibles de la comparaison. Elle dépend du degré avec lequel le contenu de l'identification peut être analysé et articulé. Rendre efficaces les analogies potentielles nécessite que la base de données du programme soit accessible pour les identifications de relations, les identifications d'objets n'ayant que peu ou pas d'utilités.

Les inférences produites peuvent être représentées par une collection de noeuds répertoriés dans le domaine cible. Ces noeuds, associés avec des drapeaux, indiquent l'origine et le mode d'obtention des données. Cela permet au programme de gérer dynamiquement la structure de correspondance par des vérifications ultérieures, suite à des confirmations ou incohérences du raisonnement. La validation ou l'invalidation de faits produits par inférence analogique, vont en conséquence agrandir ou diminuer l'étendue de la structure.

C/ FORMATION DE CONCEPTS :

1/ Un aperçu du processus d'applications d'analogies développé par CARL :

Le programme d'apprentissage CARL, développé par Burstein, construit des structures causales ou inférentielles simples dans un domaine cible en retrouvant des structures dans la mémoire d'un domaine familier et en les adaptant en utilisant un processus d'application dirigé par le haut qui préserve les liens de causalité et de temporalité explicitement spécifiés dans ces structures.

Le processus d'application essaie de former des structures dans le domaine cible sous les contraintes générales suivantes :

- les prédicats correspondants doivent être de la même classe : action, relation, étape de plan, but...
- les prédicats correspondant dans la structure du domaine de base sont en relation par des liens de causalité et de temporalité correspondant à ceux de la structure dans le domaine cible,
- les différents aspects des cas correspondants des prédicats reliés par analogie doivent, de façon consistante, être remplis par des rôles correspondants dans les 2 structures en relation.

Le programme CARL garde un enregistrement, le prototype de causalité, dans une mémoire appelée "AMAP" détaillant tout sur les correspondances développées entre les objets, rôles et prédicats. Ces enregistrements sont augmentés en cas de besoin pour inclure les correspondances trouvées ultérieurement. L'idée est que les prédicats (propriétés d'objets ou relations) appliqués dans le domaine cible sont sujets à des transformations ultérieures.

2/ Raisonnement analogique incrémental

Burstein a travaillé avec des groupes d'élèves pour étudier les différentes phases d'un raisonnement analogique. Dans les protocoles examinés, les essais pour étendre les analogies aux situations variées d'un domaine cible, avaient lieu à la fois quand les élèves répondaient aux problèmes et exemples présentés et aussi quand ils généraient des exemples dans des essais indépendants pour tester tout ce qu'il était possible dans le domaine cible.

Dans le programme CARL, chaque utilisation postérieure d'une analogie nécessite l'application d'une description des relations de causalité du domaine base au domaine cible. L'information sauvegardée dans l'"AMAP" sur les applications précédentes du prototype causal est d'abord utilisée dans un processus d'application inversée pour construire une description dans le domaine de base du nouveau problème. La recherche ou la construction de la structure de correspondance dans le domaine de base est donc dirigée par l'information préalablement contenue dans le domaine cible sur les objets sur lesquels porte l'analogie spécifiée.

Quand une analogie est spécifiée ou un nouvel exemple présenté, une session d'apprentissage se déroule en trois phases :

- déclenchement d'une application inverse du domaine cible vers le domaine source en prenant appui sur les informations contenues dans l'enregistrement du prototype de causalité correspondant,
- construction dans le domaine de base d'une description correspondant au nouveau problème,
- création de la structure de correspondance.
- une fois la structure construite, l'"AMAP" est utilisé pour appliquer la nouvelle structure causale dans le domaine cible.

L'organisation de la mémoire et le système de recherche de connaissances relatives aux objets ont été conçus en prenant exemple sur les systèmes d'indexation basés sur les objets.

Pour retrouver une variété de situation correspondant à des cas spéciaux, le processus de recherche dans la mémoire a été amélioré par l'utilisation de réseaux de discrimination basés sur les modèles de spécification de hiérarchies.

Après chaque recherche, les préconditions et les indices résultants sont ajoutés de telle sorte que les actions et les plans simples peuvent être retrouvés en réponse aux requêtes pour la réalisation du buts spécifiques, ce qui correspondant à un enrichissement de la connaissance du domaine. Une ou toutes ces formes d'indexation peuvent être utilisées dans la recherche d'une structure adaptée à appliquer, sans déclencher la recherche de nouvelles analogies.

Le programme CARL montre comment l'apprentissage analogique dans un nouveau domaine peut être accompli par un raisonnement analogique incrémental.

L'auteur du programme indique aussi que réaliser une application analogique par apprentissage demande une focalisation sur les abstractions préalablement connues dans un domaine de base. Il indique également qu'il est nécessaire de formuler des règles, à la fois pour limiter le raisonnement analogique requis pour créer les modèles initiaux de concepts dans le nouveau domaine, et pour permettre la correction incrémentale des nombreuses erreurs résultant de l'usage d'analogies.

D) ANALOGIE TRANSFORMATIONNELLE :

1) Le rôle de l'analogie dans la résolution de problème

La définition la plus commune du terme résolution de problème dérive du travail de Newell et Simon, et signifie la sélection d'une séquence d'opérateur (d'un ensemble fini préanalysé) qui transforme un état initial du problème en un état but désiré.

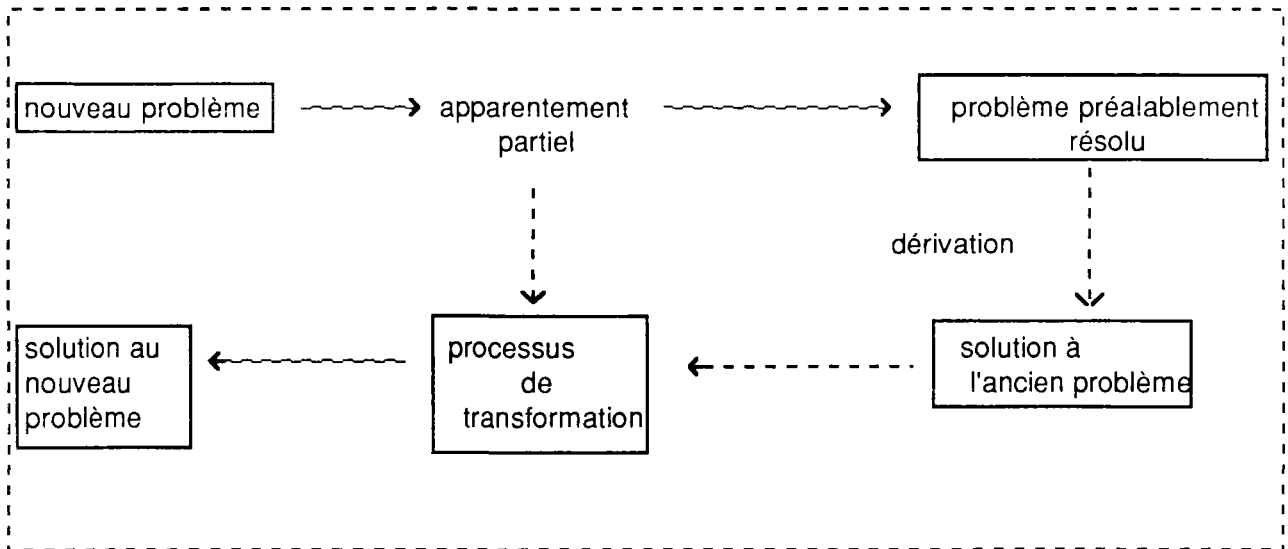
Une définition différente a été avancée par Mac DERMOTT (1967) et WILENSKY (1978, 1983) qui présente la résolution de problème comme une instanciation de plans. Pour chaque problème posé, il y a un ou plusieurs plans qui donnent les grandes lignes de la solution, qu'il faut identifier et instancier.

Une troisième approche consiste à résoudre un nouveau problème par analogie avec un problème similaire préalablement traité. C'est l'approche qui a été développée par CARBONNEL.

2) Aspects principaux de ce type d'apprentissage :

Elle constitue la méthode la plus directe pour transférer l'information sur les solutions passées au nouveau problème.

1. Recherche dans les instances des problèmes résolus celle qui s'apparente le plus à la description du problème courant.
2. Rappel de la solution du problème.
3. Transformation de la solution rappelée par un processus incrémental de réduction des différences entre ce que la solution réalise et ce qui est demandé dans le nouveau problème. Ce processus est généralement guidé par une analyse "fin-moyens" réduite à l'espace des solutions et non dans l'espace des possibles états du monde.
4. Si la transformation échoue, sélectionner une nouvelle instance candidate ou abandonner le processus d'analogie.



2) L'espace du problème :

La méthode d'analyse "moyens-fins" résout un problème en recherchant à chaque étape l'opérateur qui réduit le plus la différence entre l'état courant et le but connu. L'espace traditionnel d'un tel problème est composé de :

- l'ensemble des états possibles du problème,
- l'état initial et le but à atteindre,
- un ensemble d'opérateurs, qui, avec leurs conditions d'application transforme un état en un autre état de l'espace,
- une fonction de différence, qui calcule la différence entre deux états, et qui est généralement appliquée pour calculer la différence entre l'état courant le but final,
- une méthode pour indexer les opérateurs en fonction des différences qu'ils réduisent,
- un ensemble de contraintes que la trace de la solution doit satisfaire de l'état initial au but.

La résolution de problème dans cet espace consiste alors :

1. comparaison de l'état courant et du but
2. choix de l'opérateur qui réduit le plus la différence
3.
 - a) application de l'opérateur si possible,
 - b) sinon, sauvegarde de l'état courant, et application de la méthode au sous-problème consistant à établir les préconditions non satisfaites de l'opérateur,
 - c) quant le sous-problème est résolu, réactiver l'état sauvegardé et reprendre la méthode.

3/ L'utilisation des solutions des problèmes précédents :

a/ La première idée : la création de macro-opérateurs

Une solution étant composée d'une séquence d'opérateurs, l'idée est de les sauvegarder sous forme de macro-opérateurs indexés sur des types de problèmes caractérisés par l'état initial et le but.

L'enregistrement et la recherche de ces macro-opérateurs devient rapidement ingérable quand le nombre de problèmes traités augmente, et une solution est rarement totalement applicable d'un problème à un autre.

b/ Les deux phases de la résolution d'un problème par analogie :

La recherche d'une solution par analogie commence par le rappel des problèmes préalablement résolus et dont le contexte peut se transférer au problème en cours de considération. A ce niveau, l'utilisation de la fonction de différences comme une métrique de similitude peut servir à déterminer le problème le plus proche de celui considéré.

La seconde phase consiste à transformer l'ancienne séquence "solution" en une nouvelle satisfaisant les critères du nouveau problème. La recherche d'une transformation analogique appropriée est elle-même une procédure de résolution de problèmes, mais dans un espace différent, l'espace du problème de transformation analogique.

c/ L'espace du problème de transformation analogique :

Cet espace est défini de la manière suivante. Les états de cet espace sont constitués par les solutions potentielles dans l'espace d'origine des problèmes, c'est-à-dire, les séquences d'états et d'opérateurs de l'état initial à l'état final, avec les contraintes.

L'état initial de cet espace est la solution au problème le plus proche retrouvé lors de la première phase. Le but dans cet espace est la spécification d'une solution au nouveau problème, satisfaisant les contraintes.

d/ Les opérateurs dans l'espace de transformation :

Un opérateur dans cet espace applique une séquence "solution" dans une autre séquence potentielle "solution".

Les opérateurs les plus utilisés sont les suivants :

- | | |
|---------------------------------------|---|
| * Insertion générale : | Un nouvel opérateur est intégré à la séquence solution |
| * Exclusion générale : | Un opérateur est enlevé de la séquence solution |
| * Raccordement de séquence | Cet opérateur de transformation est utilisé quand la séquence des opérateurs du problème d'origine ne peut être appliquée à cause d'une précondition non satisfaite. Le problème d'établissement de la précondition est alors traité à part. |
| * Substitution préservant le sous-but | Un opérateur de la solution d'origine est remplacé par un autre qui réduit la même différence entre état courant et but.
Cette substitution est utile quand le premier opérateur ne peut être appliqué à cause d'une précondition insatisfaisante. |
| * Concaténation du segment terminal | Cet opérateur reprend la totalité de la séquence de la solution d'origine et complète la solution en prenant comme nouvel état initial du problème considéré l'état final de la solution d'origine. |

* Concaténation du
segment initial

Cet opérateur reprend la totalité de la solution et la complète en
prenant comme nouveau but l'état initial de la séquence d'origine.

Il existe d'autres opérateurs de transformation qui ont pour fonction de réordonner les
opérateurs, ou d'inverser une séquence complète, ou encore de substituer des paramètres.

La métrique de mesure des différences dans l'espace de transformations est une combinaison des
mesures de différences entre les états initiaux, buts, contraintes, et du degré d'applicabilité de
la solution retrouvée dans le scénario du nouveau problème.

E) LA METHODE D'ANALOGIE DERIVATIONNELLE :

1) Les limites de l'analogie transformationnelle :

En formulant des plans et en résolvant des problèmes, une quantité considérable d'information est produite en plus de la solution spécifique, comme par exemple :

- la structure des buts intermédiaires,
- la génération et le rejet ultérieur d'alternatives.

La méthode transformationnelle ignore une telle information intermédiaire se concentrant uniquement sur la séquence des opérateurs appliqués sans prendre en compte les raisons de leur sélection.

L'idée de l'analogie dérivationnelle est d'introduire dans la recherche d'une solution les étapes intermédiaires des recherches de solutions de problèmes similaires. La méthode de résolution de problème est basée sur le transfert de l'expérience passée pour les situations nouvelles de problème.

2. L'analogie dérivationnelle

Le procédé de transfert d'expérience consiste à recréer des lignes de raisonnement incluant des séquences de décision accompagnées des justifications prouvées dans la résolution de problèmes particuliers faisant appel à une analyse initiale similaire.

La dérivation analogique d'une solution à partir de solutions passées nécessite que :

- les descriptions des problèmes antérieurs et de leurs solutions soient enregistrées et indexées pour une utilisation ultérieure,
- le nouveau problème soit confronté à un grand nombre de problèmes passés potentiellement pertinents, pour trouver les plus proches, s'il y en a. Une métrique de similitude opérationnelle est nécessaire comme base de sélection des expériences passées les plus adaptées,
- la solution à un ancien problème sélectionné soit transformée pour satisfaire les conditions du nouveau problème.

a) La nécessité de préserver la dérivation des histoires.

En résolvant un problème par n'importe quel moyen, chaque étape de la solution doit être stockée, c'est-à-dire la structure des sous-buts du problème et chaque décision prise, qu'elle soit de faire une action, d'explorer de nouvelles possibilités, ou d'abandonner les plans en cours, ceci incluant :

- les alternatives considérées et rejetées,
- les raisons des prises de décision, avec les liens de dépendance à la description du problème,
- le point de départ d'une fausse piste, avec les raisons qui ont conduit à l'envisager, puis celles qui ont conduit à l'abandonner,
- dépendances des dernières décisions par rapport aux premières dans la dérivation,

- les pointeurs vers la connaissance prise en compte et reconnue utile dans la construction éventuelle de la solution.
- la solution résultante elle-même,

Dans le cas où le résolvant de problème a prouvé l'impossibilité de trouver une solution, enregistrer l'approche la plus proche, avec les raisons expliquant pourquoi un autre progrès ne peut être réalisé.

Dans le cas où la solution dépend d'affirmations inconstantes non établies dans la description du problème, enregistrer les dépendances appropriées.

b) Le procédé pour tracer des analogies par transformation dérivationnelle :

Quand un nouveau problème est rencontré et qu'il ne conduit pas lui-même à une instanciation directe d'un plan ou à une reconnaissance directe d'un cadre de solution, commencer à analyser le problème en appliquant les plans généraux ou les méthodes faibles, appropriées à la solution. Si, après avoir commencé l'analyse du problème, le processus du raisonnement fait un parallèle avec des situations passées, retrouver les traces du raisonnement complet, et utiliser le processus de transformation dérivationnelle.

Deux problèmes sont considérés comme similaires si leur analyse résulte de processus de raisonnement équivalents, au moins dans les étapes initiales. Ceci remplace la métrique de similitude plus arbitraire, requise pour la confrontation partielle parmi les descriptions de problème en traçant les analogies par transformation directe de solutions. A partir de là, les traces du raisonnement passé sont retrouvées si leur segments initiaux correspondent aux premières étapes de l'analyse du problème présent.

Les processus de raisonnement retrouvés sont alors utilisés comme des cas individuels pertinents pour générer des suppositions, et en médecine, conduire l'analyse de diagnostic. Le raisonnement à partir de cas individuels a été reconnu comme une composante importante de l'expertise, mais peu a été dit sur l'information nécessaire que chaque cas contenir.

ANNEXE 3



EVALUATION OF A HEURISTIC PROGRAM FOR SCHEDULING TREATMENT PLANT PUMPS



Texte de l'article paru dans :

"Journal of Water Resources Planning and Management",
Volume 115, du 4 juillet 1989.
Revue de l'American Society of Civil Engineers.

EVALUATION OF HEURISTIC PROGRAM FOR SCHEDULING TREATMENT PLANT PUMPS

By Pierre Lannuzel¹ and Leonard Ortolano,² Member, ASCE

ABSTRACT: PILOTE is a computer program that attempts to reproduce decisions of expert operators in scheduling outlet pumps at a water treatment plant near Paris, France. The program integrates both mathematical models and heuristics, i.e., rules of thumb used by operators to select configurations of outlet pumps. The models include statistical procedures for predicting hourly water consumption and a hydraulic simulation for estimating hour-by-hour elevation changes at regulating reservoirs and at the elevated reservoir receiving the pumped flows. The heuristics used by operators are based on factors related to water quantity and quality and to operation and maintenance costs. An evaluation of the program showed that the statistical forecasting procedures performed well, but the simulation model consistently overestimated the volumes pumped. The heuristics included in the program accurately represented rules used by expert operators, except for rules used to rank-order alternative pumping strategies. The evaluation provided an agenda of research for refining PILOTE so that it can be used routinely.

INTRODUCTION

This paper elaborates on the development and evaluation of PILOTE, a computer program to assist in the real-time operation of a water treatment plant located in a suburb of Paris. The plant is owned by the Syndicat des Eaux d'Ile de France and operated by a private firm, the Compagnie Générale des Eaux (CGE). PILOTE is designed to assist plant operators in scheduling the hour-by-hour use of the plant's ten outlet (or lift) pumps over a 24-hr period. Scheduling decisions are complex because of uncertainties in water consumption, and they are important because the plant's ability to supply consumers can be jeopardized if the decisions are made poorly.

Operators currently schedule pumps based on experience. Novice operators sometimes have difficulty meeting water demand targets while maintaining high water quality and low operating costs. PILOTE was developed to improve overall plant performance by making the reasoning used by expert operators available to novices. The computer program relies on heuristics elicited from expert operators, and it is intended to reproduce the performance of these operators in scheduling pumps.

Although PILOTE exhibits some characteristics of an expert system, its reasoning is not "transparent" (Maher 1986); the term "heuristic program" provides a more suitable description (Levin et al. 1986). As explained by Pearl (1984), heuristics are methods and criteria that provide a simple means for deciding among numerous courses of action. Heuristics do not necessarily yield "optimal" results, but they typically yield good solutions to prob-

¹Researcher, Ecole Nationale des Ponts et Chaussées, CERGRENE, Nonsy-le-Grand, France.

²UPS Foundation Prof. of Civ. Engrg., Dept. of Civ. Engrg., Stanford Univ., Stanford, CA 94305, on sabbatical leave at CERGRENE during 1987-88 academic year.

Note. Discussion open until December 1, 1989. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on August 4, 1988. This paper is part of the *Journal of Water Resources Planning and Management*, Vol. 115, No. 4, July, 1989. ©ASCE, ISSN 0733-9496/89/0004-0457/\$1.00 + \$1.50 per page. Paper No. 23674.

r = coefficient;
 α = coefficient, dimensionless;
 SD = Secchi disk depth, m;
 SW = amount of energy absorbed by individual model layer, kcal/m²/h;
 Z = depth below water surface, m and
 β = percent of incident light absorbed lake surface, dimensionless

lems within a reasonable time. For a discussion of heuristics and their use as an approach to problem solving, see Pearl (1984) and Barr and Feigenbaum (1981).

PILOTE was developed at the Centre d'Enseignement et de Recherche pour la Gestion des Ressources Naturelles et de l'Environnement (CER-GRNE), a natural resources research laboratory that is part of the Ecole Nationale des Ponts et Chaussées. An ad hoc "management committee" that included heads of several CGE departments followed the program's development. The committee viewed PILOTE as an important element in the broader CGE effort to improve performance of water treatment plants by the increased use of computer-based decision support systems. The technical director of the water treatment plant was a member of the committee and played the role of "super expert" when individual expert operators used different heuristics as a basis for running the facility.

OUTLET PUMPING SYSTEM

The sequence of units at the water treatment plant consists of intake pumps, flocculation tanks, settling basins, filtration tanks, ozonation units, regulating reservoirs, and outlet pumps. There are three groups of constant speed outlet pumps, and they are operated from two regulating reservoirs designated as R_1 and R_2 , respectively (see Fig. 1). Water can be drawn from R_1 by two small pumps and six medium pumps. There are two large pumps operating from reservoir R_2 .

The regulating reservoirs serve as a buffer between the other units that make up the plant and an elevated reservoir that feeds into the network, distributing water to consumers. A configuration of outlet pumps lifting water from the regulating reservoirs to the elevated reservoir is represented by the vector (i, j, k) , where: i = number of large pumps in operation, $i = 0, 1, 2$; j = number of medium pumps, $j = 0, 1, \dots, 6$; and k = number of small pumps, $k = 0, 1, 2$. Because each outlet pump operates at a constant speed, a 24-hr schedule for the pumps, herein termed a "complete strategy," is given by 24 three-element vectors (one per hour). The term "partial strat-

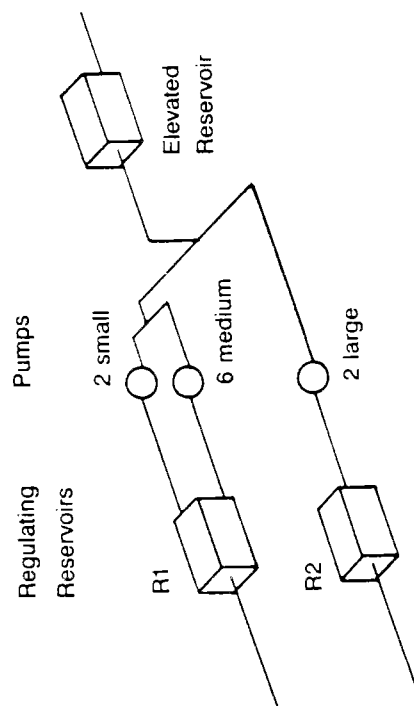


FIG. 1. Outlet Pumping System at Plant

egy" refers to a set of three-element vectors describing pump usage for less than 24 hours.

The number of conceivable complete strategies is enormous. For any hour, there are three possible values for the small (or large) pumps (i.e., 0, 1, or 2 pumps) and seven possible values for medium pumps. Consequently, there are 63 conceivable configurations ($3 \times 7 \times 3$) for each hour and 63^{24} possible complete strategies.

MODULAR STRUCTURE OF PILOTE

In deciding on pump configurations to be used during their work shifts, expert operators do not search all conceivable partial strategies. Rather, they use rules-of-thumb based on experience to make their decisions. Inasmuch as PILOTE aims to reproduce the performance of expert operators, it limits the space of strategies investigated by means of heuristics provided by the operators. In addition, it describes alternative complete strategies using multiple decision criteria employed by expert operators. PILOTE does not optimize a multiple-criteria evaluation function subject to constraints. The identification of a suitable evaluation function is a topic for research.

Fig. 2 shows PILOTE's principal modules and indicates how both heuristics and mathematical models fit into the program. The "context" module allows for the input of background information such as day of the week, season, and the beginning and end of several planning periods of less than 24 hours. It also allows for specifying conditions under which the plant is operating, such as "normal," reservoir R_1 closed for repair, or three medium pumps out of service. Rules-of-thumb used by operators change with different contexts, and PILOTE uses this module to activate the appropriate sets of heuristics. In this paper, only the heuristics for normal operating conditions are presented; PILOTE can treat a limited number of other contexts.

Once the context is established, PILOTE uses a statistical model to predict hourly consumption by water users over the 24-hr period under consideration. Based on anticipated total consumption, PILOTE calculates the flow rate sequence for the input pumps. Operators give a high priority to keeping input flows constant because changes in flow lead to problems with water quality (e.g., breakup of materials clogged on filtration beds and resuspension of materials deposited at the bottom of settling tanks). However, depending on the expected water consumption, it may not be possible to run the plant with only a single input flow. Complications arise because of conditions in the contract that CGE has with Electricité de France (EDF), the electricity supply organization: outlet pumps cannot be run during certain hours of the day. If expected water demand is very high, the fixed capacities of the three reservoirs may make it impossible to meet the needs of consumers with a single input flow rate. PILOTE checks on this using standard hydraulic calculations. If a single flow is not feasible, PILOTE checks if two input flows during 24 hours will suffice. The computation continues in this fashion as necessary. Often, however, a single input flow is adequate.

The cycles shown in Fig. 2 are part of a "generate-test-select" loop that is initiated by specifying a planning period and "parents." The latter are defined as partial strategies that are feasible at the end of one period and that continue into the next period.

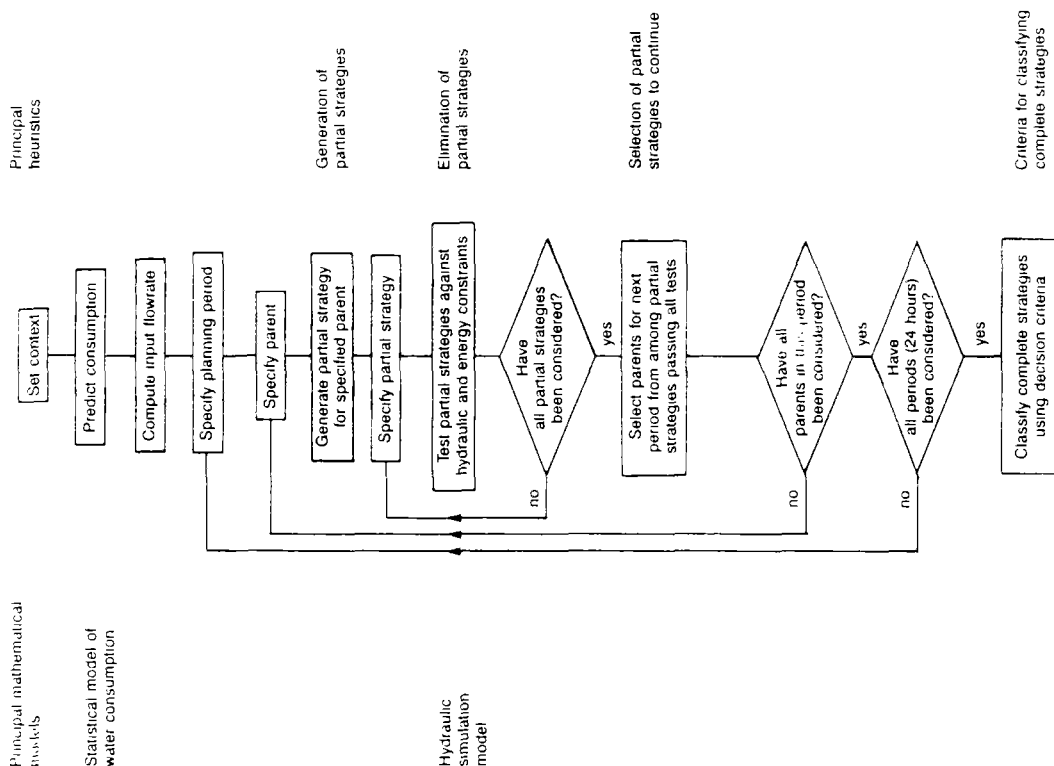


FIG. 2. Structure of PILOTE

Consider the first planning period (~24 hours). Rules based on procedures followed by operators are used to eliminate a subset of the 63 conceivable outlet-pump configurations available for each hour. Partial strategies not thereby eliminated are said to be "generated." They are analyzed individually in the "test" module. A hydraulic simulation model is used in testing, and it takes the initial levels of the R_1 , R_2 , and elevated reservoirs as a starting point. For any partial strategy, the simulation model calculates the hourly changes in elevation of all reservoirs, as well as energy consumption and hydraulic pressures at various points. Heuristics provided by expert operators are then used to eliminate partial strategies that violate conditions, such as the re-

quirement for a minimum volume in the elevated reservoir consistent with maintaining a reliable water supply. Partial strategies that remain after the test module are pruned further in the "select" module. In this case, decision criteria employed by operators are used to determine which partial strategies will be extended into the next planning period. The generate-test-select loop continues until planning periods constituting 24 hours have been considered.

The final module in PILOTE presents all complete strategies identified, as well as ratings for the decision criteria used by operators. PILOTE does not pick a strategy. The function of PILOTE is to help improve plant performance, not to make final choices.

The next two sections indicate how the mathematical models used in PILOTE were evaluated. This is followed by a description and evaluation of PILOTE's heuristics.

MODELS OF WATER CONSUMPTION

PILOTE includes two alternative models for predicting hourly water consumption, each of which is based on an unpublished study done in 1987 by Penet. He developed a linear regression model for predicting total daily consumption of the water treatment plant's output based on season, day of the week, consumption on previous days, and the estimated temperature of the day in question. For any predicted daily consumption, hourly consumption is calculated using empirically derived coefficients for weekdays, holidays, etc. in a given season. A regression model based primarily on Penet's analysis, but using parameters estimated from a more extensive data set, is included in PILOTE. Penet also made preliminary observations about the potential value of an auto-regressive integrated moving average (ARIMA) model. Taking Penet's work as a starting point, an ARIMA model using historic water consumption data was developed for inclusion in PILOTE.

The comparative evaluation of the ARIMA and linear regression models involved both a traditional analysis of forecasting errors and an investigation of how such errors could influence outlet-pump scheduling decisions. Since the traditional analysis was based on well-known procedures, e.g., an examination of root-mean-square errors, it is not discussed extensively. The traditional error analysis indicated that both models performed reasonably well and that neither was superior. Although prediction errors as high as 20% were sometimes observed, errors in consecutive hours typically balanced each other. The result was that cumulative errors over any three consecutive hours were always less than 10%, and cumulative errors for any planning period were always less than 5%. The traditional error analysis was based on data for five weeks in two seasons.

The analysis of how prediction errors can influence pump scheduling decisions required establishing a benchmark condition for comparative purposes. The benchmark was defined as ratings for criteria used in selecting outlet pump strategies when PILOTE is run retrospectively with observed values of water consumption. These decision criteria ratings were compared with those obtained using predicted water consumption for the same 24 in period. All other aspects of PILOTE were identical in this analysis.

As an example, consider a case involving only the period from 11:00 A.M. to 6:00 P.M. on February 15, 1988. Table 1 contains PILOTE's output at the end of the generate-test-select loop for the 11:00 A.M. to 6:00 P.M. period. All

TABLE 1. Error Analysis Based on Ratings for Decision Criteria (Partial Strategies Generated by PIOTE for February 15, 1988, 11:00 A.M.–6:00 P.M.)

Configuration elevated reservoir (1)	Security elevated reservoir (2)	Security EDF (3)	Water pressure (4)	Distance from elevated reservoir target (5)	Distance from regulating reservoir target (6)	Configuration changes/ period (7)	Electricity cost (francs/m ³) (8)
(a) Benchmark Condition—Observed Water Consumption							
111	Good	Good	Good	Very poor	Very poor	0	11.1
112	Good	Good	Good	Very poor	Very poor	0	11.3
120	Good	Good	Good	Poor	Very poor	0	11.2
121	Good	Good	Good	Average	Average	0	11.5
(b) Consumption Predicted with Regression Model							
111	Average	Good	Good	Very poor	Very poor	0	11.1
112	Good	Good	Good	Very poor	Very poor	0	11.3
120	Good	Good	Good	Very poor	Very poor	0	11.2
121	Good	Good	Good	Poor	Average	0	11.5
(c) Consumption Predicted with ARIMA Model							
111	Average	Good	Good	Very poor	Very poor	0	11.1
112	Good	Good	Good	Very poor	Very poor	0	11.3
120	Good	Good	Good	Very poor	Very poor	0	11.2
121	Good	Good	Good	Poor	Average	0	11.5

runs of PIOTE indicated that only four partial strategies were to be considered in subsequent planning periods: (1,1,1), (1,1,2), (1,2,0), and (1,2,2). Each of these candidate configurations was to be initiated at 11:00 A.M. and continued until 6:00 P.M.

Table 1 lists ratings for criteria considered by expert operators in deciding among alternative strategies. These criteria are described in the following order of importance to the operators.

1. Security-elevated reservoir—lower limits for water levels at the elevated reservoir consistent with meeting basic requirements of water users.
2. Security-EDF—consistency with EDF contract restrictions regarding electricity usage during different hours.
3. Water pressure—upper bounds on pressure consistent with maintaining the integrity of the distribution network.
4. Distance from elevated reservoir target—difference between water level at the elevated reservoir and operator's target elevation at a particular time.
5. Distance from regulating reservoir targets—difference between water levels in regulating reservoirs and operator's target elevations at a particular time.
6. Configuration changes/period—number of changes in outlet-pump configurations per planning period.
7. Unit electricity cost—cost of electricity at plant per cubic meter of water lifted by outlet pumps

These decision criteria were first identified in interviews with operators, and a follow-up questionnaire was used to define ordinal measures for each criterion. (The last two criteria are exceptions: they are measured using numerical values.) The measurement process is clarified by considering "security-elevated reservoir." Operators indicated that 3.5 m was the minimum acceptable water level value at the elevated reservoir consistent with security considerations. PIOTE calculates hourly water levels at the elevated reservoir for each partial strategy run through the simulation model, and the

difference between the lowest hourly elevation and the security level (3.5 m) provides the basis for measuring "security-elevated reservoir." In response to questionnaires, operators specified numerical cut-off values used to categorize the aforementioned elevation difference as "good" (i.e., no security problems anticipated), "average," or "poor." Using these cut-offs, an ordinal rating is assigned for each partial strategy considered in the simulation model.

The measurement procedures for security-EDF and water pressure are similar to security-elevated reservoir. The security-EDF rating is based on the minimum difference between hourly electricity consumption and the maximum allowable consumption (as specified in the contract with EDF) for the corresponding hours. The basis for the water pressure criterion score is the minimum difference between hourly total pressure head at a specific location and the maximum allowable pressure specified by operators. The measurement procedures for distance from elevated, (regulating) reservoir targets also use operator-specified cut-offs, but they pertain only to differences between elevations and targets at the ends of periods.

Values underlined in Table 1 indicate criteria and strategies for which errors in the models for predicting consumption are potentially problematic. With only three exceptions, criteria had the same ratings in the benchmark case and in runs using the forecasting models. Exceptions concerned the security-elevated reservoir criterion for configurations (1,1,1) and the distance from the elevated reservoir target for configurations (1,2,0) and (1,2,1). Despite the differences, the relative ranking of the partial strategies, using a lexicographic ordering method described later, was the same.

The analysis in Table 1 was repeated for nine planning periods in February 1988, and the results were all similar to those listed in Table 1. The only criteria for which results from the forecasting models yielded ratings different from the benchmark cases were security-elevated reservoir and distance from elevated reservoir target. This occurs because the computation of water level at the elevated reservoir is sensitive to errors in predicted water consumption. The forecasting models must be refined, since security-elevated reservoir is the single most important decision criterion used by expert operators in selecting an outlet pumping strategy. Errors in the security-elevated reservoir criterion ratings can influence the selection of strategies in unacceptable ways.

HYDRAULIC SIMULATION MODEL

PIOTE's hydraulic simulation model estimates hourly changes in water levels of the R_1 , R_2 , and elevated reservoirs for any partial strategy. Principal relationships include curves of flow rate versus total pressure head for each pump type and an empirical equation derived by Lequenne (unpublished report, 1987) for calculating total pressure head at the exit from the outlet pumps. Lequenne's relation estimates head as a function of elevations in the three reservoirs, hourly consumption by water users, and outlet pump configuration. Lequenne also derived empirical equations for calculating total electric power consumption as a function of flow rate lifted by an outlet pump configuration. The simulation model also includes the law of conservation of mass for flows entering and leaving the reservoirs.

Simulation model inputs include: initial water elevations at the R_1 , R_2 , and

elevated reservoirs, total pressure head at the exits from the outlet pumps at the start of a planning period, input pump flow rate, predicted hourly water consumption, and the partial strategy under consideration. The model yields hourly values for the entire planning period of water elevations at all reservoirs, total pressure heads at the outlet pumps, and energy consumed by the plant.

Evaluation studies for the simulation model were based on strategies implemented by operators during nine planning periods in February 1988. For each period, the model was run retrospectively using the implemented strategy and the actual not predicted, water consumption for the period. The model's outputs were compared with corresponding observed values.

There was excellent agreement between observed and predicted values for water pressures at outlet pumps (differences of $\pm 1\%$). However, for energy consumption, the model yielded values 1–5% less than corresponding observed values. The percentage underestimate was generally constant for each hour of a planning period, and this suggests the possibility of improving model performance by reestimating coefficients in the equations used for predicting energy consumption.

A significant shortcoming of the simulation model concerns volumes pumped from the regulating reservoirs to the elevated reservoir. In a typical hour, the volume lifted was overestimated by 3–6%. By the end of a nine-hour planning period, the overestimate in total volume pumped was sometimes as high as 50%. As shown later, errors of this magnitude have significant adverse effects on PIOTE's ability to accurately represent the decision making of operators. Because the errors involve a consistent overestimation of volume pumped, it may be possible to improve the simulation by recalibrating some of the empirical equations using a more complete set of observed data.

HEURISTICS FOR STRATEGY GENERATION AND ELIMINATION

Reduction to 16 Configurations

Interviews were conducted with several expert operators to elicit the heuristics they use in deciding which strategies to examine under normal plant operating conditions. These interviews were unstructured and were conducted while operators were working at the plant. Direct observation of the configurations operators considered in the course of making their decisions supplemented the interview data. Results indicated that operators use factors related to electricity consumption, water quality, and pump efficiency to eliminate all but 16 of the 63 configurations possible for any hour.

EDF Contract

For each season, the contract between the Compagnie Générale des Eaux and Electricité de France specifies an hour-by-hour schedule of maximum allowable energy consumption at the water treatment plant. Operators take these contract provisions seriously because of the substantial financial penalties associated with violations. Operators do not normally consider configurations that simultaneously use two large pumps and three or more medium pumps because they are likely to violate limits in the EDF contract.

Water Quality

Under normal conditions, operators pump water from reservoirs R_1 and R_2 simultaneously in order to minimize changes in input-pump flow rate and

thereby maintain high water quality. If only one of the two regulating reservoirs is in service, there is an increased chance of having to change the flow-rate sequence for the input pumps during a 24-hr period. The reason relates to maximum energy use provisions in the EDF contract that require outlet pumps to be shut down during selected hours. When all outlet pumps are off, flows from the plant accumulate in the regulating reservoirs. If only one regulating reservoir is in service, only about half the potential storage for flow regulation is available, and it may be necessary to change the input flow rate twice, to decrease input just prior to the time when outlet pumps are shut down and to increase input when the pumps are brought back into service.

Pump Efficiency

Past experience has shown that configurations involving three or more medium pumps will have a relatively low efficiency. With these configurations, the velocity and turbulence in regulating reservoir R_1 increase substantially due to suction at pump intakes. The result is a notable drop in efficiency.

Further Reductions Using Approximate Mass Balance

Other heuristics concern water elevations. Operators routinely judge the acceptability of partial strategies by the status of the R_1 , R_2 , and elevated reservoirs at selected hours. Based on experience, they have established water levels that they consider desirable at particular times; these are termed "target elevations" herein. Questionnaires were distributed to operators to obtain target elevations and values above and below each target that are acceptable under normal operating conditions.

A procedure based on operator-specified target elevations was used to reduce the number of partial strategies examined. Consider, for example, a situation in which the acceptable elevations at the elevated reservoir are $5.5 \text{ m} \pm 0.5 \text{ m}$ at the end of a planning period that runs from 11:00 A.M. to 6:00 P.M. A mass balance is used to compute the volume at the elevated reservoir at 6:00 P.M., and this volume is compared to upper and lower bounds on volumes that an outlet-pump configuration must lift to be acceptable. (Because there is a correspondence between water elevation and volume stored at a reservoir, acceptable elevations are equivalent to a corresponding set of volumes.) In this instance, the acceptable elevations of $5.5 \text{ m} \pm 0.5 \text{ m}$ are expanded to $5.5 \text{ m} \pm 1 \text{ m}$, and the target volumes associated with the expanded range of values are determined. Use of an expanded range of volumes compensates for the very approximate way in which volumes pumped are calculated.

The elevated reservoir volume at 6:00 P.M. is computed as the volume at the elevated reservoir at 11:00 A.M. plus the volume lifted during the seven-hour planning period minus the volume used by consumers during that period. At this point in PIOTE, both the 11:00 A.M. volume at the elevated reservoir and predicted hourly values of water consumption are known. The 6:00 P.M. volume (and water level) at the elevated reservoir can be calculated once the volume lifted has been estimated.

The principal approximations in this mass balance are in calculating volumes lifted using each strategy. To simplify this example, assume that once a configuration is selected at 11:00 A.M., it is used until 6:00 P.M. Thus,

the volume lifted by a configuration is seven (hours) times the hourly volume pumped. The latter is calculated by assuming that total head at the outlet pumps is constant (at roughly the average historically observed value for a particular season and planning period). Although the head depends on the configuration, this assumption is reasonable, since the range of acceptable water levels at the elevated reservoir is two times the range specified by operators. Assuming that head is constant, flows from various pumps are obtained from curves of flow rate versus head.

Using these simplifications, the volume at the elevated reservoir can be computed for each of the 16 configurations. Partial strategies that yield volumes outside acceptable bounds are eliminated.

Minimizing Shifts in Configuration

Even using all previously noted heuristics, there are still n^3 possible complete strategies ($n \approx 16$). A heuristic related to minimizing changes in configurations further shrinks the space of conceivable strategies.

Expert operators routinely keep the number of configurations used in a work shift low, and in normal circumstances, they almost never use more than three configurations per shift. Excessive switching of outlet pumps is a sign of poor performance for several reasons. First, there is the increased wear and tear that results from turning valves and from switching pumps on and off. This increases maintenance costs and may disrupt normal plant operations. A second reason concerns water hammer and the associated risk of having pipes burst. (Operators sometimes observe substantial short-term variations in pressure within pipes downstream of outlet pumps when a pump configuration is changed.) Another motivation for minimizing changes in outlet pumps concerns switching that may be required in input flow rate and consequent adverse effects on water quality. Regulating reservoir levels often fluctuate widely when outlet pumps are modified, and input flows may need to be changed to meet target elevations at R_1 and R_2 .

The practices of expert operators regarding switching configurations are translated into constraints within PILOTE. For example, under normal operating conditions in winter months, PILOTE uses at most three configurations from 8:00 P.M.–6:00 A.M., at most two from 11:00 A.M.–6:00 P.M., and only one from 6:00 A.M.–9:00 A.M. Under these circumstances, the maximum number of complete strategies that might be considered is reduced from n^3 to n^6 ($n \approx 16$). Constraints on switching are different in other seasons, but reductions in possible strategies are nonetheless substantial.

Heuristics for Partial Strategy Elimination

Application of the given strategy-generation heuristics yields a set of partial strategies which are analyzed further using the hydraulic simulation model. Each such strategy is simulated to determine the hour-by-hour values of energy consumption, water pressure, and water elevations in the reservoirs.

One group of heuristics for eliminating strategies during a simulation concerns elevations. If during any hour, a reservoir level goes above the reservoir's maximum usable height or below an operator-specified minimum consistent with security, the associated strategy is eliminated. In addition, strategies are dropped if they yield water levels (or equivalent volumes) at the R_1 , R_2 , and elevated reservoirs that are outside ranges of acceptable target values specified by operators. Because elevations estimated using the sim-

ulation model are more accurate than those calculated using the approximate mass balance analysis, the simulated elevations are compared to the original, not expanded, ranges of acceptable elevations specified by operators.

A second group of strategy elimination heuristics concerns energy usage allowed by the EDF contract. A strategy is dropped if it causes estimated energy consumption to become close to the maximum allowable value for any hour.

EVALUATION OF HEURISTICS FOR GENERATING AND TESTING STRATEGIES

Rules for generating and testing partial strategies were evaluated in the context of real-time plant operations for nine planning periods in February 1988. At the beginning of a period, operators on duty were asked to specify the partial strategy to be used, as well as the strategies they considered as alternatives to the selected strategy. After the planning period was over, PILOTE was run using the observed, not predicted, water consumption. Partial strategies generated by PILOTE for eventual consideration in the hydraulic simulation model were then compared with strategies chosen by the operators. Observed, not predicted, water consumption was used in PILOTE to avoid having errors in the prediction model influence the validation of the heuristics.

For five of the nine periods, the strategies generated by PILOTE included all strategies considered by operators in charge. This is the result expected if the strategy generation and testing heuristics were correct. The five instances were all afternoon periods in which the pump strategy selection decision was relatively simple. Typically, the operators considered only two or three strategies and PILOTE generated only four or five strategies. In all cases, the pumping programs involved only one configuration per period.

Table 2 summarizes results for the four periods (all 8:00 P.M.–6:00 A.M.) in which PILOTE failed to generate all strategies judged worthwhile by operators in charge. Although PILOTE generated many more strategies than the operators, it missed one strategy per period. This shortcoming is explained by errors in the hydraulic simulation model. As noted previously, the simulation model consistently overestimated the volume lifted to the elevated reservoir by the outlet pumps. For every instance in Table 2 in which PILOTE failed to generate a strategy considered worthwhile by operators, the operators' strategy involved more pumped flow than any strategy considered by PILOTE. By overestimating the volume pumped, PILOTE incorrectly calculated that too much water would be lifted. The volume lifted using each strategy in the last column of Table 2 was so great that it caused the upper limit for the target water level at the elevated reservoir to be crossed, thus the strategy was eliminated. The same reasoning explains why strategy 112/121 was dropped in the testing phase using the simulation model for February 25, 1988.

These results indicate that the generation and testing heuristics in PILOTE were consistent with the behavior of operators in a real-time setting for the periods included in the evaluation study. The results also show that problems with the hydraulic simulation model have adverse effects on PILOTE's ability to accurately represent the operators' decision making.

TABLE 3. Ratings for Partial Strategies Subject to Simulation (PILOTE Runs Using Actual Water Consumption: 8:00 P.M. 2/22/88-6:00 A.M. 2/23/88)

Partial strategy number ^a (1)	Security elevated reservoir (2)	Security EDF (3)	Water pressure (4)	Distance from elevated reservoir target (5)	Distance from regulating reservoir target (6)	Configuration changes period (7)	Final d_p test (banks m) ^b (8)
1 (111)	Poor	Good	Good	Very poor	Very poor	0	10.7
2 (112)	Average	Good	Good	Very poor	Very poor	0	10.4
3 (120)	Average	Good	Good	Average	Poor	0	10.4
4 (121)	Average	Good	Average	Good	Good	0	10.6
5 (122)	Good	Good	Good	Very poor	Very poor	1	10.3
6 (111)/(112)	Poor	Good	Good	Very poor	Very poor	1	10.4
7 (111)/(120)	Poor	Good	Good	Poor	Very poor	1	10.4
8 (111)/(121)	Poor	Good	Good	Poor	Very poor	1	10.4
9 (112)/(120)	Average	Good	Good	Average	Poor	1	10.4
10 (112)/(121)	Average	Good	Average	Average	Very poor	1	10.4
11 (120)/(112)	Average	Good	Good	Good	Average	1	10.5
12 (120)/(121)	Average	Good	Good	Good	Very poor	1	10.4
13 (121)/(112)	Average	Good	Good	Good	Very poor	1	10.4
14 (121)/(120)	Average	Good	Good	Good ^d	Average	1	10.5

^aIf one vector is shown, it represents a constant configuration from 8:00 P.M. 6:00 A.M. Other strategies have one change first configuration for 8:00 P.M. 1:00 A.M. and second configuration for 1:00 A.M. 6:00 A.M.
^bUnderlines indicate the tie scores in the application of the lexicographic ordering method.
^cStrategies eliminated during tests made in simulation run.
^dUsing lexicographic ordering, strategy 14 is selected as the one partial strategy for a particular initial parent to continue into the next planning period.

applies in all periods except the first, uses operator-specified decision criteria to eliminate all but the most promising partial strategy derived (within a period) from each parent node.

Suppose that at the start of the second period there are m parent nodes, each of which has an associated set of initial reservoir elevations based on simulation model results for the last hour of the first period. Each parent node is a starting point for a maximum of $16p$ partial strategies to be generated during the second period, where p = the maximum number of different configurations (within the period) per partial strategy. The value of p is between one and three, depending on the period and season. Based on experience, the number of partial strategies generated and remaining acceptable at the end of a period is typically less than 15 per parent.

PILOTE uses a lexicographic ordering method (Voogd 1983) to select the one partial strategy per parent from one period to continue into the next period. It begins by considering the top-priority decision criterion: security-elevated reservoir. If one strategy surpasses all others in terms of security-elevated reservoir, it is selected. If there are tie scores among competing strategies, the decision is based on the second priority criterion: security-EDF. This process continues until there are no ties. Table 3 illustrates the method using criteria ratings for each strategy generated from one of m parents existing at the start of the second period in an analysis of the 24 hours beginning at 11:00 A.M. on February 22, 1988. The final decision is based on the fourth most important criterion: distance from the elevated reservoir target. Strategy number 14 is chosen as the one parent (of the 13 possible) for starting the third period. Using a similar computation for each of the other ($m - 1$) parents available at the start of the second period yields at most m parent nodes for the start of the third period.

TABLE 2. Evaluation of Heuristics for Strategy Generation and Testing, Planning Period from 8:00 P.M.-6:00 A.M. Following Morning (* Indicates Strategy Implemented by Operator)

Date (1)	Strategies considered by operators (2)	Number of strategies generated by PILOTE ^a (3)	Strategies considered by operators and not generated by PILOTE (4)
2/22/88	120 ^b 121 *120/121 ^c 120/122	14	120/122
2/23/88	120 121 120/121 *120/122	13	120/122
2/24/88	112/121 *120/121	12	120/121
2/25/88	120 *112/121 ^d 120/121	13	120/121

^aAll PILOTE outputs were generated a posteriori using observed water consumption.
^bWhen a single configuration is shown, it indicates that there is no change in configuration during the period.
^cWhen two configurations are shown, the first is used until 1:00 A.M., and the second is used thereafter.
^dThe configuration 112/121 was generated by PILOTE, but it was eliminated during testing within the hydraulic simulation.

HEURISTICS FOR DETERMINING PARTIAL STRATEGIES TO CONTINUE

As shown in Fig. 2, going from a partial strategy to a complete strategy requires repeated use of the generate-simulate-test loop until consecutive periods covering 24 hours are analyzed. Each partial strategy that remains acceptable at the end of a period can be viewed as a "parent node" in the portion of the tree of possible strategies generated during the next period. (There is only one parent node at the start of the first period.)

For each parent existing at the start of a period, it is necessary to generate and test all partial strategies. In conducting tests, the number of required simulation runs increases substantially as the number of parents grows. The resulting increase in computation time is problematic, since PILOTE is intended for use in real time. Even with all heuristics introduced, it would require more time than operators have available in order to simulate all feasible complete strategies on microcomputers used at the plant. Computation time can be kept within available bounds by limiting the number of parents. The approach used to restrict the number of parents at the start of a new period is the only instance within PILOTE in which strategies are eliminated based on procedures not employed by expert operators. This method, which

Operators do not use heuristics to determine which strategies to continue into the next planning period because they only make decisions for their assigned work shifts. It was possible, however, to investigate whether use of lexicographic ordering might be defensible by examining the multicriteria decision process operators use in selecting a strategy for a particular period. This was done by having each of several operators solve case study strategy-ranking problems. For each case, an operator was presented with six partial strategies representing reasonable options for use in a given planning period. Each partial strategy had been simulated using the same conditions of water consumption, initial reservoir elevations, and initial water pressure at the outlet pumps, and results from the simulation runs were given to the operators. Each operator was asked to rank-order the six alternative strategies that constituted a case study problem using the procedures they would normally follow in selecting an outlet-pump strategy. In some instances, operators were asked to "think out loud" and thereby describe (to an interviewer) the reasoning used in ranking strategies.

Five case study strategy-ranking decision problems were given to three operators, and the same five cases were also run through PILOTE to yield decisions based on lexicographic ordering. Results showed clearly that the decision process used by operators is more complex than lexicographic ordering.

Operators began each case study by considering their top priority criterion: security-elevated reservoir. Typically, two or three partial strategies were dropped immediately because of low security-elevated reservoir ratings. In considering other criteria for the strategies that remained, operators made deliberate trade-offs. For example, a strategy with an average level of security but no changes in configuration might be preferred over a strategy with an excellent level of security but two changes in configuration. In each instance, trade-offs were made in an ad hoc, informal way. The lowest priority criterion, unit cost of electricity, never received serious consideration. Lexicographic ordering yielded the same final decision as the operators in only two of the five case studies. In many instances, different operators did not make the same final decisions.

Because the case studies showed that operators' procedures for selecting an outlet-pump strategy are poorly understood and not well represented by lexicographic ordering, no attempt was made to rank the complete strategies yielded by PILOTE. Instead, each complete strategy is presented with corresponding ratings for the decision criteria.

Neither the technical director of the treatment plant nor the CGE management committee was prepared to have the technical director play the role of "super expert" to develop a procedure for ranking alternative strategies. The development of a multicriteria ranking function is a topic requiring further research.

CONCLUSIONS

The evaluation of PILOTE provided an agenda for research needed to bring the program to the level where it accurately represents outlet-pump

scheduling decisions of expert operators. The simulation model needs improvement, especially the portion estimating the volume lifted by a configuration. Refinements must also be made in the models for predicting water use because consumption forecasting errors can significantly influence values of the security-elevated reservoir decision criterion. In addition, work is needed to develop a procedure that represents the way operators use multiple criteria to rank alternative strategies.

After reviewing the evaluation of PILOTE, the CGE management committee suggested additional refinements. They wish to expand the number of operating contexts to include all conditions representing routine, predictable variations from normal operating conditions. They also wish to extend PILOTE so that it can be re-run automatically within a planning period to exploit reservoir water level data collected routinely on an hourly basis.

Although PILOTE is not yet ready to be used in routine plant operations the work on PILOTE has been useful to CGE in several ways. The heuristics employed in operating the water treatment plant have been clearly identified, and the criteria used in selecting outlet-pump strategies have been articulated carefully for the first time. In addition, the models for predicting water consumption can be used (in a restricted context) to provide operators with training in the value of statistical prediction techniques. This is significant, since operators currently do not have a systematic way of predicting consumption. Finally, the current version of PILOTE is useful in training inexperienced plant operators. Future refinements in PILOTE will allow CGE to integrate it into its broader program of using computer-based decision aids to improve plant operations.

ACKNOWLEDGMENTS

The writers are grateful to Catherine Perman, Department of Civil Engineering, Stanford University, for helpful comments on an early draft of this paper. Suggestions from the two anonymous reviewers are also appreciated.

APPENDIX. REFERENCES

- Barr, A., and Feigenbaum, E., eds. (1981) *The handbook of artificial intelligence*, Vol. 1, Addison-Wesley, Reading, Mass.
- Levin, R. I., Rubin, D. S., and Simson, J. P. (1986) *Quantitative approaches to management*, 6th Ed., McGraw-Hill, New York, N.Y.
- Maier, M. L. (1987) "Expert system components," *Expert systems for civil engineering: Technology and applications*, M. L. Maier, ed., ASCE, New York, N.Y.
- Pearl, J. (1984) *Heuristics: Intelligent search strategies for computer problem solving*, Addison-Wesley, Reading, Mass.
- Voozid, H. (1983) *Multicriteria evaluation for urban and regional planning*, Pion Limited, London, U.K.